2D PHYSICAL PROPERTIES IN THE RESPONSES OF THE MACAQUE
INFEROTEMPORAL CORTEX

DOCTORAL THESIS
2005

TAMÁS TOMPA

SUPERVISOR: PROF GYÖRGY BENEDÉK

UNIVERSITY OF SZEGED
ALBERT SZENT-GYÖRGYI
MEDICAL AND PHARMACEUTICAL CENTER
FACULTY OF MEDICINE
Publications, providing the basis of the thesis


IV. Tamas Tompa, Zoltan Chadaide, Laura Lenti, Gabor Csifcsak, Gyula Kovacs and Gyorgy Benedek
Invariances of shape-processing for reduced surface cues: how IT neurons and psychophysics correlate in the macaque
# TABLE OF CONTENTS

1. Introduction ............................................................................................................... 1
2. General Materials and Methods ............................................................................... 5
3. Effect of colour on IT neuronal responses of the macaque .................................... 10
4. Effect of surface details of stimuli on the IT neuronal responses of the macaque 18
5. Possible category-related features in the responses of IT neurons ....................... 34
6. Conclusion .............................................................................................................. 49
# Detailed Table of Contents

1. Introduction .................................................................................................................. 1

2. General Materials and Methods .................................................................................. 5
   a. Subjects .................................................................................................................. 5
   b. Surgery .................................................................................................................. 5
   c. Apparatus .............................................................................................................. 7
   d. Stimuli ................................................................................................................... 7
   e. Stimulus sequence and behavioural paradigms ..................................................... 8
      i. Single cell recording ......................................................................................... 8
      ii. Single cell recording protocol ........................................................................ 9
   f. Data analysis .......................................................................................................... 9
      i. Responsivity and selectivity ............................................................................. 9
      ii. Comparing the selectivities ............................................................................ 9
      iii. Comparing the responses ............................................................................. 9

3. Effect of colour on IT neuronal responses of the macaque ....................................... 10
   a. Introduction ......................................................................................................... 10
   b. Specific methods ................................................................................................. 11
      i. Stimuli ............................................................................................................... 11
      ii. Subjects .......................................................................................................... 11
   c. Results ................................................................................................................. 12
   d. Discussion .......................................................................................................... 14

4. Effect of surface details of stimuli on the IT neuronal responses of the macaque .... 18
   a. Specific methods ................................................................................................. 18
i. Stimuli

ii. Subjects

b. Results

i. Effects of colour, texture, shading and internal contour removal on responsiveness
   1. BLD
   2. DLD
   3. SIL
   4. LD
ii. Effects of colour, texture, shading and internal contour removal on shape selectivity
   1. BLD-DLD
   2. DLD-SIL
   3. Overall
   4. LD
   5. LD-COL
iii. Response latencies
iv. Analysis of different response intervals

c. Discussion

i. Texture and shading
ii. Silhouettes
iii. Contrast
iv. LD
   1. LD as hunting stimuli
v. Effect of practice
5. Possible category-related features in the responses of IT neurons ................. 33
   a. Introduction.................................................................................................. 33
   b. Specific methods........................................................................................ 34
      i. Subjects.................................................................................................. 34
      ii. Statistics................................................................................................ 34
         1. Cluster analysis (Ward method).................................................... 34
         2. Factor analysis ............................................................................. 35
         3. Multidimensional scaling (MDS)................................................... 35
         4. Physical features of the images.................................................. 36
         5. Sparseness index, selectivity...................................................... 36
         6. Latency times.................................................................................. 37
   c. Results.......................................................................................................... 38
      i. General findings................................................................................ 38
      ii. Results of cluster analysis............................................................ 38
      iii. Results of the factor analysis....................................................... 40
      iv. Physical features........................................................................... 42
      v. Sparseness, selectivity index...................................................... 43
   d. Discussion.................................................................................................... 44

6. Conclusion........................................................................................................ 48

7. References........................................................................................................ 49
2D PHYSICAL PROPERTIES IN THE RESPONSES OF THE MACAQUE INFEROTEMPORAL CORTEX

1. INTRODUCTION

In the course of the evolution visual systems first evolved to detect changes in the environment and to keep control over the movements of the animal itself. (Bruce et al 1996, Goodale-Westwood 2004). Vision as “perception” (“sight” of things in the environment) evolved relatively late in the natural history. It is not difficult to see however that this innovation increased greatly the animals’ potential for flexible, adaptive behavior by enabling them to carry out complex cognitive operations on mental representations of the world. This historical and functional division is reflected in the functional anatomy of the higher order visual system of the primates. The dichotomy was first described by Ungerleider and Mischkin (1982). As Goodale and Millner (1992) later proposed, a dorsal “action” stream would project from early visual areas to the posterior parietal cortex providing flexible visual control of motor acts; and a ventral “perceptual” pathway would project from the same early areas to the temporal lobe, providing a rich and detailed representation of the environment for object recognition and pattern identification. This theory was later supported by several physiological, anatomical and even histological studies (as an extensive review: Caminiti et al 1996).

The ventral pathway is composed of a set of posterior cortical areas extending from the primary visual cortex (V1) to the inferotemporal (IT ) cortex. It can be characterised by a hierarchical architecture in which neurons in higher areas code for progressively more complex representations by pooling information from lower areas. Neurons in V1 code relatively simple features, such as local contours and colours, whereas neurons in IT fire in
response to whole complex objects. (see for example Lee 2003, Tanaka 1996, Logothetis and Sheinberg 1996) In this hierarchical organization the size of the receptive fields (RF, the area of the visual field to which a neuron responds) changes characteristically. Whereas in V1 neurons have small RFs thus being able to provide precise spatial information about the position and details of the stimulus in the visual field (Lee 2003), in the TE, where neurons receive massive convergent inputs from lower areas, the RF’s are large (Gross et al 1972, Gross 1994). This feature can underlie our capacity to recognize objects independently of their size and position. (Tanaka 1996, Logothetis and Sheinberg 1996) On the other hand, large receptive fields mean the loss of spatial information and thus lead to difficulties, for example in representing several objects at a time. (Zoccolan et al 2005). The retinotopic organization becomes coarser along the pathway. No sign of retinotopy has been reported in TE (Tanaka, 1997), although in humans a kind of center-periphery organization has been revealed in corresponding areas by fMRI experiments. (Levy 2001).

Of course this somewhat oversimplified view of a hierarchical cortical processing cannot be held entirely true, especially in the light of recent advances in perceptual science.

The early cortical areas do not seem any more to be a mere early,¹ simple processing stages which get relatively rough information, processing it in its course and giving it forward to more sophisticated, higher areas. In the emerging dynamic vision of the cortex, the “early” areas can have an important role in several stages of the visual processing. Several integrated models are proposed on the basis of recent data. One of them regards V1 and V2 as an “active black-board” for the rest of visual cortical areas (Bullier & Nowak 1995). Another one (Lee 2003) regards them as high resolution buffer, relying on the fact that the actually the only part in the visual pathways which has a high enough resolution to provide fine-grained representation of the environment are the early visual areas with their 0.5-1 degree RFs. Indeed it was shown that higher-order visual processes, such as spatial attention (Ito and Gilbert 1999) or object attention (Roelfsema et al 1998) affect V1. There are illusory contour responses (Lee and Nguyen 2001) and also “shape-from-shading” pop-out responses (Lee at al 2002) reported at this level.

Thus, whilst there might be some basic hierarchy in the cortical processing of the information, in the sense that some areas are relying on the output of others, it is also true that this processing is not serial at all. Depending on response onset times it is clear that

¹ “…a mere filter bank in the first stage of processing…”
information is processed simultaneously in multiple cortical areas and feedback and feed-forward streams are shaping the process at any time. (Bullier and Nowak 1995, Lee 2003, Deco and Lee 2004)

On the other hand it is also shown, that it is theoretically possible for the higher areas to accomplish their task (i.e. categorizing) from a simply serial flow of information (Delorme and Thorpe 2001, VanRullen 2005), and there are some who argue that most of the information carried by the sensory neurons is available in the beginning of the spike-trains (Tovee et al 1993), which would make the feed-back mechanism redundant.

The anterior part of the inferotemporal cortex (IT) is thought to be essential for object recognition. Multiple streams of evidence support this idea. First, cortical ablation studies have demonstrated that lesions of the IT produce selective impairments in object recognition (Dean, 1976; Logothetis and Sheinberg, 1996). Second, IT neurons respond in a highly selective manner to complex stimuli from objects differing in shape, colour and/or texture (for reviews see Logothetis and Sheinberg, 1996; Tanaka, 1996). The shape selectivity of these neurons parallels the invariances of object perception in several ways: shape preference of IT neurons is largely unaffected by changes in the position and size of an object (Schwartz et al., 1983; but see Ito et al.,1995, Cox et al 2005), by the defining cue (Sáry et al., 1993; Tanaka et al 2001) and by partial occlusion (Kovács et al. 1995a).

Despite the considerable amount of work done in the field it is still unclear what features are important for the TE neurons and how they shape the responses.

To reveal the role of two-dimensional surface features in the visual processing in the object area TE, we planned a series of experiments in non-human primates.

Our everyday experience teaches us that object recognition is to a large extent independent of a range of changes in the retinal image, i.e. the change (and also reduction) in the surface detail of the object. We do not find more difficult to recognize the same object when seen in reality, on a coloured or greyscale photograph, or even depicted as a line drawing. This phenomenon has been widely used in the arts and by professional illustrators (e.g. pictograms). Indeed, in a human psychophysical experiment, Biederman and Ju (1988) found that the naming latencies of masked objects presented as coloured photographs or as line drawings were essentially the same. A series of experiments revealed no benefits for chromatic over achromatic representations (Ostergaard and Davidoff, 1985) or over line drawing representations (Davidoff and Ostergaard, 1988) in different classification tasks.
This suggests that surface characteristics such as colour, texture and shading play only a secondary role in object recognition once contour information is available. This finding is in line with edge-based theories of object recognition (Biederman, 1987; Grossberg and Mingolla, 1985; Ullman, 1989).

Hayward et al (1999) compared the abilities of human subjects to recognise silhouettes and shaded images of objects rotated in depth. They found that humans use the 3D representation for object recognition and that silhouettes provide only partial 3D information, due to the lack of shading. This suggests, that the interpretation of 3D structures of objects is enhanced by shading and internal contours (Cavanagh, 1991).

In the following experiments, we systematically examined whether the shape selectivity of IT neurons is dependent on changes in retinal input caused by variations of the surface attributes of the presented objects. We employed twenty standard stimulus. In the first experiment we examined the effect of the removal of the information carried by the colour. Then each object inside its occluding contours was systematically reduced. We removed the texture and shading, keeping the inner contours. The contrast polarity was varied at this stage. Then, the internal contours were also removed, leaving merely a silhouette of the object. During the experiments, we recorded the single-cell activity of certain IT neurons in awake, fixating monkeys. For each individual neuron, from a standard set of 20 coloured objects we first identified 2 object to which the neuron responded vigorously (effective stimuli) and 2 to which it did not respond (non-effective stimuli). Next, we compared the responses of the neurons to these 4 objects presented under the progressively reduced conditions. This procedure shares some similarities with the step-by-step stimulus reduction paradigm of Tanaka et al (1991), during which a 3D object is gradually reduced by removal of its colour, texture, shading, contours and object-parts, which allows determination of the critical features for the neurons in anaesthetised animals. However, there is a conceptual difference between that systematic stimulus reduction method and our method. Instead of first identifying an object that a particular neuron responds to and then reducing it to determine a feature that is still essential for maximal activation of the neuron, we used both the original objects and their reduced variants as stimuli in order to be able to compare the behaviour of the neurons at a population level.
2. GENERAL MATERIAL & METHODS

a) Subjects

Three adult macaque monkeys (two Macaca mulatta, monkey C and Ch and one Macaca nemestrina, monkey K) were used as subjects. To achieve appropriate behavioural prerequisites for the experiments we used the „controlled water access“ paradigm. For 20 hours preceding the experimental sessions the monkeys were deprived of water. After the daily experimental sessions, the animals received supplementary water, vitamins, fruits and vegetables as necessary and had access to dry food ad libitum. Recordings were generally made for 2-3 hours a day, 4-5 times a week. During one session, each monkey typically consumed 200-300 ml of water or fruit juice. The weight of the animals was checked regularly and was kept at 90% of the normal body-weight. Special attention was paid to the animals’ general condition, with frequent checks on their body weight, fur and excrement. Training or recording sessions were interrupted for one month in every 2-3 months.

b) Surgery

Before the surgery, the animals were adapted to the laboratory and to the primate chair. A scleral search coil was implanted into one eye, according to the procedures described by Judge et al (1980). At the same time, a stainless steel peg was cemented to the skull for head fixation purposes. The head of monkey K was fixed by the reversible method developed by Pigarev et al (1997). A recording

Figure 1 Reconstruction of the recording site (monkey Ch and K).
The chamber was next implanted over the anterior dorsolateral part of the skull of the animals, following the procedures of Vogels (1999b). The position of the recording chamber was determined with the help of magnetic resonance and CT images taken before the operation. The centre of the recording chamber was situated 17 mm anterior to the auditory meatus and 24 mm lateral to the sagittal midline over the left hemisphere in monkey C, and 17 mm anterior to the auditory meatus and 23 mm lateral to the sagittal midline over the right hemisphere in monkey K. These coordinates for monkey Ch were 16 mm anterior and 23 mm lateral. The chamber of monkey C was tilted 6 degrees inward, while that of monkey K and Ch were positioned vertically. Recording chambers were implanted over both hemispheres in monkey C, but all recordings were made in the chamber positioned over the left hemisphere. (See reconstructed recording tracks on Figure 1.) All surgical procedures were carried out under full anaesthesia and under aseptic conditions. Anaesthesia was initiated with an i.m. injection of ketamine (Calypsol; 8 mg/kg) and atropine (0.05 mg/kg). An endotracheal tube was placed into the trachea and anaesthesia was maintained with a mixture of N₂O and O₂ in a ratio of 2:1. An intravenous line was inserted for continuous access and additional fentanyl (i.v. 2-4 mg/kg) was given whenever necessary. Before the surgical procedure, a preventive dose of antibiotic was given (i.v. Augmentin, 500 mg amoxycillin and 100 mg clavulanic acid). The same doses of antibiotics were given i.v. on the first 5 postoperative days. The incision was infiltrated with local anaesthetic (Procaine). Nalbuphin and non-steroidal anti-inflammatory drugs were administered to the animals postoperatively. Arterial oxygen saturation, expired CO₂ level, heart rate and rectal temperature were monitored continuously throughout the surgery; and kept within normal limits.

At the end of the recording sessions, several penetrations were made in monkey C and K with stainless steel wires under ketamine anaesthesia. The monkies were then sacrificed with an overdose of Nembutal and perfused with fixative. Recording sites were reconstructed by identifying the tracks of the last few penetrations in coronal brain sections (100 µm) stained with cresyl violet. Monkey Ch is still being used for ongoing experiments (Chadaide et al 2004). All procedures conformed to the guidelines of the NIH for the care and use of laboratory animals and were approved by the Ethical Committee of the University of Szeged.
c) Apparatus

During the recording sessions, the monkey sat in a custom-made primate chair with its head fixed. A standard 17-inch monitor (refresh rate = 74 Hz) was placed in front of the animal 57 cm from the eye. A PC recorded eye movements (200 Hz sampling rate), delivered the reward and controlled the animals' behaviour. Other computers presented stimuli and collected electrophysiological data.

Sterile tungsten electrodes (FHC, parylene-coated with an impedance of 1.0-2.0 MΩ), held by a Narishige hydraulic microdrive, were used for single-cell recordings. Signals were amplified, frequency-filtered and fed into the recording PC, audio monitor and oscilloscope. Single-cell discrimination was performed with an amplitude window discriminator for monkey C and with a spike separator system (SPS-8701, Real Time Waveform Discriminator System, Prospect, Australia) for monkey K and Ch. The background luminance in the experimental room was kept constant at a level < 1 cd/m².

d) Stimuli

Stimuli were presented on a uniform grey background square (side: 18 degrees; luminance = 8 cd/m²) positioned in the center of the screen. A set of chromatic stimuli (COL) composed of 20 figures was used (Figure 2). Half of the figures were simple geometrical shapes filled with a coloured, textured pattern, created by a commercial image processing software. The stimuli occupied the same area (6 x 5 degrees) and had an average luminance of 7.9 cd/m² (SD=5.6 cd/m²). The other half were chromatic images of natural and artificial objects (occupying the central 10 x 7 degrees of the screen, with an average luminance of 4.8 cd/m² (SD=3.0 cd/m²), chosen randomly from the image pool of the laboratory. Stimuli were presented centrally during the fixation of a small blue fixation spot (0.1 degree radius and 5.5 cd/m² luminance) that remained on screen throughout the trial. We made transformations of the stimuli as described in the appropriate sections below.
e) Stimulus sequence and behavioural paradigms

i) Single-cell recording

A simple fixation paradigm was used during the single-cell recording sessions. Initially, the screen was black. A trial started with the presentation of the fixation spot. If the animal foveated the fixation spot, the uniform grey background pattern was presented for 500 msec, after which the stimulus appeared for another 500 msec. Animals were rewarded for maintaining the fixation within a 0.5 x 0.5 degree square window until the stimulus offset. If they left the fixation window earlier than the stimulus offset, the trial was considered ‘aborted’ and excluded from further analysis. To associate the reward and stimuli, but not the fixation spot, reinforcement was given immediately after the stimulus offset, while the fixation spot remained onscreen for a variable time (100 to 300 msec). The inter-trial interval was 1000 msec.

In the object discrimination task, the stimulus offset was followed by the appearance of two circular red targets (0.45 deg radius and 2 cd/m² luminance) flanking the fixation spot at a distance of 7.5 degrees to the right and left. The monkey was trained to make a saccadic
eye movement to the left or to the right target after the presentation of each individual COL stimuli. These stimuli were classified into 2 groups of 10 figures such that the discrimination task could not be solved by the presence or absence of one particular feature in the images.

**ii) Single-cell recording protocol**
We searched for single cells by presenting our standard set of 20 COL stimuli. Once a cell was isolated and found to be responsive to at least one of the COL stimuli, it was tested further. To test stimulus selectivity, we ran tests by presenting 4 objects, 2 eliciting larger firing rates of the particular neuron and 2 less effective stimuli, determined by auditory feedback and upon inspection of the peristimulus time histograms. Each of the 4 objects was then presented as COL and under the transformed stimulus conditions. Each stimulus condition was presented at least 10 times in an interleaved fashion.

**f) Data analysis**
Off-line spike counts were computed trialwise with a 500 msec bin, starting 50 msec after the stimulus onset. Net responses were calculated by trialwise subtraction of the neural activity during a fixation period of the same duration as the stimulus time window, but just preceding the stimulus onset. Analysis of variance (ANOVA, Kirk, 1968) was used to test the significance of the responses to the stimuli and the significance of shape selectivity. Tests were classified as significant if the corresponding type I error was smaller than 0.05.

**i) Responsivity and selectivity** To determine the responsiveness and selectivity of the individual neurons a two-way ANOVA was performed on the neural data with the stimuli and the time period of the firing activity (before vs. after stimulus onset) as factors. A cell was considered responsive if the main effect of the responses was significant, and it was considered selective if the responses to the stimuli differed significantly. We performed this analysis on each neuron for each condition individually.

**ii) Comparing the selectivities** To compare the selectivities of the neurons for the different conditions we used another two way ANOVA with the conditions (COL and BW) and the stimuli (4) as factors. To be able to compare the selectivity on a population level, for each neuron and condition we ranked the responses given to the 4 objects, according to their net responses under the COL condition.

**iii) Comparing the responses** The responses to the different conditions were
compared by generating responsivity indices (RIs). For each cell, we subtracted the average net firing rates in response to the preferred stimulus under the transformed condition from the average net response to the preferred stimulus under the COL conditions and divided this difference by the sum of the two responses. 

$$SI = \frac{(R_{max})_{COL} - (R_{max})_{TR}}{(R_{max})_{COL} + (R_{max})_{TR}}$$

where $$(R_{max})_{COL}$$ is the response of the given neuron for the preferred stimulus in the COL condition and $$(R_{max})_{TR}$$ is the response of the given neuron for the preferred stimulus in the transformed condition. The closer the value of RI to 0, the more similar the responses of the neuron are to the two conditions.

To compare the selectivities of the cells under the two conditions, we determined a selectivity index (SI). For each cell and each stimulus condition, we subtracted the average net firing rate in response to the least preferred stimulus (i.e. rank 4) from the average net response to the preferred stimulus (i.e. rank 1) under the same stimulus conditions and divided this difference by the sum of the two responses:

$$SI = \frac{R_{max} - R_{min}}{R_{max} + R_{min}}$$

where $$R_{max}$$ is the maximal and $$R_{min}$$ is the minimal response to a stimulus from the stimulus set, respectively. The closer this index is to “1”, the more selective is the cell, i.e., the bigger the difference between a “preferred” and a “non-preferred” stimulus.

To compare the time courses of the spike trains in response to two conditions in selected cases, we used a paired t-test on a fine-resolution population histogram generated from the averaged responses to the first ranked stimuli. The areas under the curves were compared.

The response onset latency was calculated by using the “Poisson spike train analysis” of Legéndy and Salcman (1985). In this analysis, for each cell and stimulus, the trialwise median of the onset times of the first activations was used as latency.

3. EFFECT OF COLOUR ON IT NEURONAL RESPONSES OF THE MACAQUE

a) Introduction

The discrimination of colours is an important, though not mandatory feature of the ability to recognize objects (Gegenfurtner & Sharpe 1999). As the last unimodal part of the ventral visual pathway (Mishkin & Ungerleider 1982) the IT is connected to areas proved to possess
colour-sensitive cells (Saleem et al 2000, Tamura & Tanaka 2001) and its output (ie. amygdala, rhinal cortex etc.) requires a modestly sophisticated level of processing, which might include colour information., (Cheng et al 1998 Saleem & Tanaka 1996) The crucial role of the IT in the recognition of objects (Gross CG 1994, Tanaka 2000) makes it likely to comprise an important part of the colour-processing system.

Livingstone & Hubel (1987, 1988) have demonstrated that colour and form (altogether with motion) are processed separately in early visual areas. It is still an open question as to whether these types of information are combined or remain separate in the representation of objects in higher areas. The possibility of the existence of independent channels for shape and colour in the ventral visual pathway (Komatsu 1998) requires a closer investigation of the effects of colour on the IT neuronal responses.


On the other hand, colour is not always of crucial importance in object recognition (Biederman & Ju 1988, Delorme et al 2000). This has also been demonstrated by a few studies as reflected in the activity of IT cells (Nakamura et al 1994, Booth & Rolls 1998, Vogels 1999b).

We have conducted a study in which we tested shape-selective IT neurons for chromatic and achromatic figures. We compared the shape preferences of the cells in the two conditions and found rather similar response characteristics.

b) Specific Methods

i) Stimuli
We created the achromatic grayscale versions (BW) of the images by removing their colour. The stimuli were isoluminant with the COL versions (mean= 8.1 cd/m², SD= 1.16 cd/m², median= 8 cd/m²). These shapes contain surface detail only in the form of texture, luminance gradients (shading) and inner contours.

ii) Subjects
The subjects in these experiments were Monkey K and Monkey Ch.

c) Results
We recorded the responses of 56 IT neurons responsive for either COL or BW shapes. 45 of them were responsive under both conditions (48 for COL and 51 for BW). Those responsive were tested further for both conditions. On comparison of the averaged firing patterns for the two conditions in the case of the preferred stimuli, no statistically significant difference was observed in the fixating or in the discriminating animal. (Fig 3)

Most of the neurons that were responsive and selective in the COL condition remained responsive and selective in the BW condition (33 of 39, 84.6%). An example of a unit tested with chromatic and achromatic versions of the same 4 shapes is presented in Figure 4. This neuron responded in a highly selective manner to the 4 COL shapes and responded to them, retaining its overall shape selectivity under achromatic conditions.
To determine the selectivity of the neurons first, for each neuron and condition we ranked the 4 shapes according to their net responses in the COL condition. Next, we calculated the average net firing rate separately for each unit in the COL and BW conditions as a function of the shape rank. Figure 5a presents the relationship between figure rank and average net response in the COL and BW conditions. The slope of the curve is similar in the two conditions (ANOVA, interaction between colour and shapes; \(n = 33\), current effect: \(F(3, 96) = 1.3668, p = 0.2576\) NS), indicating that colour removal did not affect the overall shape selectivity. A cell-by-cell analysis demonstrated that most of the neurons (26/33, 79%, Table 2.) displayed the same shape preference for the chromatic and achromatic shapes (ANOVA, interaction of rendering condition and shape rank on the net firing rate for each cell individually, NS).

The mean RI for the 2 conditions was 0.042815 (\(n = 33\), min: -0.395918, max: 0.805790, SD:0.254525) which suggests a highly similar responsivity of the population to the coloured and achromatic stimuli. (Figure 5b) The SI did not differ for the 2 conditions (t-test: \(p = 0.191931\) NS). This value strengthens the result obtained by the ANOVA above. The latencies of the responses for the 2 conditions did not differ (COL: 141.27 ms, BW: 144.56 ms \(p = 0.758308\) NS). However the latency data for the two monkeys in both conditions revealed a statistically significant difference (in COL: 119.63 ms vs. 156.88 ms \(p = 0.000069\) and in BW: 119.57 ms vs. 163.45 ms \(p = 0.000305\)) which is to be explained in the future.
d) Discussion

Colour is a salient cue for the segmentation of shapes (Nothdurft, 1993). Whether or not colour also improves object recognition is widely disputed in the literature and is known to depend on several factors (Biederman & Ju, 1988, Wurm et al, 1993). However, it is certain that, under everyday circumstances, humans (and also monkeys (Vogels, 1999a)) are able to identify the shape of an object when it is presented as a coloured or an achromatic image. Our finding, that IT neurons respond in a highly similar fashion to chromatic and achromatic representations of the same shapes accords with the hypothesis that the IT plays an important role in the very flexible process of object recognition. Shape preferences of IT neurons are largely invariant on change of the position, retinal size (Ito et al 1995) and defining visual cue (Sáry et al 1993) or surface cue (Kovács et al 2003) of the stimulus, properties paralleling invariances of visual perception. Some IT neurons also proved to be selective for colour (Desimone et al. 1984, Tanaka et al., 1991, Komatsu et al. 1992, Edwards et al. 2003) and this selectivity is independent of shape selectivity (Komatsu & Ideura, 1993). These findings made the importance of IT in colour processing controversial in the literature. Some ablation studies detected colour-discrimination deficits after lesions or cooling of the IT cortex or parts of it (Buckley et al 1997, Horel 1994). Heywood et al (1998) found that a lesion in the anterior IT abolished the colour-induced pupillary reaction in contrast with the ablation of area v4. Earlier Heywood et al (1995) demonstrated that bilateral lesions to the IT disrupted hue-discrimination ability of the macaque.

Figure 5ab A) Comparison of the selectivities for the COL and BW conditions, B) Responsivity index (see in Methods) for the COL and Bw conditions
Edwards et al (2000) reported that IT and STS (superior temporal sulcus) face-selective neurons responded differentially to faces presented in greyscale or in false colours. Tamura & Tanaka (2001) observed different colour sensitivity characteristics of the subregions of the IT, the ventral part having the stronger colour sensitivity. Other studies revealed that the IT contained many colour-selective neurones (Komatsu et al 1992), and these neurones exhibited a selectivity for hue and saturation (Komatsu 1993). Some IT neurons display a sustained activity in working memory-related task for particular colours. (Fuster & Jervey 1982). Functional brain imaging studies have also suggested a role for the IT in colour processing: in a PET study, Takechi et al (1997) found that The IT is activated by a colour discrimination task. Edwards et al. (2003) recently reported a strong colour dependence of the response of a majority (70%) of temporal cortex cells to complex stimuli in one macaque. On the other hand 80% of the inferotemporal neurones studied by Nakamura et al (1994) did not change their response characteristics after the removal of colour and their stimulus-selective properties likewise remained unchanged. Booth & Rolls (1998) found that a majority of IT neurons responded similarly to chromatic and achromatic versions of the same stimulus. Vogels also observed in categorizing monkeys that, although the elimination of colour had an effect on the response rates, many IT neurons were invariant as concerns the chromatic – achromatic stimulus change and on average the stimulus preference was largely invariant as regards this stimulus transformation (Vogels 1999b, p1254 and fig13E). (although in his Discussion he stresses the importance of colour in the IT).

Our results detailed above support the notion that colour information is not a crucial cue in the processing of object information for most of the IT cells. Not only did a majority of the recorded cells retain their response amplitude (mean firing rate) for their preferred stimuli, but the overall selectivity also remained the same. The population response curves for the preferred stimuli in the 2 conditions overlap fully (Figure 3)

The mean response rates and the peak response amplitude in response to the COL and BW stimuli did not differ in either animal. However, a difference was noted between the 2 animals. While the peak response amplitudes were the same, in the case of the discriminating animal the response had a more sustained character (Figure 3). The differences observed might be attributed to the difference in the tasks, since one of the monkeys was only involved in a fixation task, whereas the other actively discriminated the stimuli. Similar attention-related differences have been described elsewhere (Super et al 2001). Despite this difference,
since the animals were presented with the same stimulus set, we believe that our argumentation about coding COL and BW stimuli in the IT are not affected by this difference.

However, our results do not indicate that the IT does not have a role in colour processing. It might be the case that colour-sensitive cells are clustered in the cortex. This possibility is suggested by Komatsu et al (1992) whose recording technique includes a search for colour-sensitive cortical areas and the implantation of chronic guiding tubes for electrodes. A clustering of colour-sensitive cells in the cerebral cortex would not be unique: blobs are thought to be such clustered colour-sensitive cells in v1 (Livingstone & Hubel 1988). Similarly, Nakamura et al (1994) found only a few (3 of 53) neurones which proved to be clearly colour-selective. An accepted model of the columnar organization of the IT (Fujita et al 1992, Tanaka 1996, 1997, 2000) states that it is the combined activity of cells organized in feature-representing columns that represents an object in its entirety. In this case, the question arises of whether there are separate cells in the IT cortex to process colour information or whether the represented features all have their own preferred colour. Our method of recording shape-selective neurons and testing them with a limited set of chromatic and achromatic stimuli allows verification of the presence of a colour-independent shape-processing mechanism in the IT (although it provides virtually no information on the nature of the colour processing itself.) It appears plausible that colour processing in the IT occurs independently of shape processing, either through separate cell populations or through a different coding mechanism.

Our results are in clear contradiction with those of a recently published study (Edwards et al 2003) in which it was found that when complex stimuli were presented in a rapid serial visual manner (duration: 111 ms, inytertrial interval: 0 ms, RSV presentation), a majority of the recorded temporal cells were sensitive to the changes of the colour information, and the selectivity for these complex shapes was also affected by removing colour. Edwards et al concluded that the temporal cortex exhibits a strong colour sensitivity, with a priority over the shape selectivity.

The differences from our findings seems to be mainly due to the differences in methodology. The rapid presentation without a time gap between the stimuli poses the question of the possibility of a masking effect (Kovács 1995), which can cause differences in the results.
Further, the fact that “the spike counts were not corrected by subtraction of the background firing rate from the stimulus response” (p1252) (which is due to the RSV presentation) can affect the results; we always worked with net responses, avoiding the effect of the possibly changing background noise.

The high proportion of face-selective cells suggests that the recording sites were located on the upper part of the IT (the lower bank of the STS) (Fig. 1 of Edwards et al strengthens this), while we tended to record cells from deeper parts of the IT.

Although the number of recorded cells is indicative of a robust result, the use of one subject only leaves open the possibility of idiosyncratic recording. Our recordings along a considerable length of the IT in 2 monkeys furnish trustworthy results.

The setting of the response latency of those cells that do not clearly exhibit a response to 100 ms as Edwards et al (2003) did, again poses the problem of contaminating the results of the latency-aligned population histograms.

Edwards et al (2003) calculated a “colour sensitivity” index, which is analogous with our “selectivity index” and indicated a very strong sensitivity to colour in the studied population. Our selectivity index had a value of 0.13 which indicates a strong insensitivity to colour in the population of cells that we examined. This suggests that the two cell populations might be different.

Overall, our results suggest that a colour independent shape processing mechanism does exist in the IT which can play a role in the invariant recognition of simple and complex, chromatic and achromatic shapes.

4. **Effect of surface details of stimuli on the IT neuronal responses of the macaque**

a) **Specific Methods**

   i) **Stimuli**

   Four different stimulus transformations of these 20 images were carried out. To remove all texture and shading information, we generated line drawings with a uniform surface brighter or darker than the background. These images retained their inner contours and the contrast
between the inner object surfaces and the background with the two opposite polarities intact. (1) Bright line drawings (BLDs) were obtained by removing the internal texture, shading and colour from the images and replacing them with a uniform white (39 cd/m² luminance and 66% contrast, as compared to the background). Black lines revealed the outer and main inner contours of the images, which had a thickness of 3 arc min, a contrast of 88%, as compared to the background grey, and a luminance of 0.5 cd/m². These main contours were determined at the main discontinuities at the minima of negative curvatures and at the large narrowings of the shapes without minima of negative curvature. This resulted in mostly convex object segments (Biederman, 1987, 1995), delineating the main parts of the objects. Lines, bordering these main parts and falling inside the shapes are defined as inner contours of the stimuli. (2) Dark line drawings (DLDs) were made in a similar way to the BLDs, but the inner surface of the objects was filled a uniform dark-grey (1.5 cd/m² luminance and 68% contrast). The inner black lines were identical to those in the BLDs. (3) Line drawings (LDs) were generated by filling the inner surface of the objects with the background uniform grey and by removing all contrast from the images, except at the outer and inner contours, which were drawn with lines identical to those in the BLDs and DLDs. (4) Silhouettes (SILs) were obtained by filling the objects with the uniform dark-grey used in the DLDs and removing all surface detail, leaving only the occluding contours and the contrast present in the image. The DLDs and SILs differ only in the presence/absence of the black lines corresponding to the inner contours of the objects.

ii) Subjects
Monkey C and K served as subjects for this study.

b) Results
The present study is based on 149 neurons that proved to be visually responsive and selective for the chromatic versions of the objects (67 and 82 neurons in monkey C and K, respectively). The remaining neurons are not considered further here. Table I lists the numbers of cells, recorded under each stimulus condition.
i) **Effects of colour, texture, shading and internal contour removal on responsiveness**

1) **BLD**

For 90 neurons, we tested how the removal of internal texture and shading cues and their replacement with a uniform surface brighter than the background (BLD) affects the neural responses. Figure 6a presents examples of the stimuli and the responses of a typical IT neuron for the COL and 3 reduced stimulus conditions. This neuron responded vigorously to the chromatic versions of stimuli #20 and #17 in Figure 2. These responses were not significantly different from those observed under the BLD condition (Scheffe's post hoc analysis, p>0.7 for each stimulus). Furthermore, the shape selectivity was also preserved in the responses under each condition (ANOVA, interaction between rendering condition and stimuli; (F(3,74)=0.28; NS)).

Most of the recorded neurons (74; 82%) that were responsive and selective under chromatic conditions were also responsive after the removal of texture and shading under the BLD condition. At a population level, the neurons, however, responded less strongly to a given object when its texture and shading were removed than to the chromatic version of the same image. Figure 7a shows the distribution of the RI for the COL-BLD comparison. The median RI was 0.17 (1\textsuperscript{st} quartile, 0.05; 3\textsuperscript{rd} quartile, 0.37; n=90), indicating that overall the response strength under the texture-removed BLD condition
Figure 6 Two IT cells responding to the surface reduced conditions.
was approximately three-fourths of that under the COL condition, a small, but significant change (Wilcoxon matched pair test; T=566; p<0.001; n=90).

2) DLD
For 77 neurons, we tested how change of the sign of the contrast between the object and the background alters the neural responses. As can be seen from Figure 6a, the neuronal responses and selectivity were not altered when the objects were brighter or darker than the background surface: this neuron responded in a similar fashion to the objects when presented as COL, BLD or DLD (Scheffe’s post hoc analysis, p>0.8 for each object). The RI of this cell for the COL-BLD and for the COL-DLD comparisons was -0.11 and -0.02, respectively, showing similar responses. Figure 7b depicts the distribution of the RI of the recorded neuron population for COL and DLD with a median of 0.27 (1\textsuperscript{st} quartile, 0.017; 3\textsuperscript{rd} quartile, 0.65; n=77) suggesting a somewhat, but not significantly larger response reduction under the DLD than in the BLD condition (means ± standard errors for the BLD and DLD indexes are: 0.33±0.05 and 0.46±0.08, respectively; Wilcoxon matched pair test for Rank 1 objects in BLD and DLD conditions: T=702; N.S.; n=77). However, the response strengths under the BLD and DLD conditions correlated well (Spearman R=0.58; p<0.001), suggesting no significant differences for objects with opposite signs of contrast.

There were neurons for both the BLD and DLD conditions (12; 13 % and 20; 26 %) with RI over 0.8, suggesting that, at least for some cells, removal of internal shading did affect response rates.

3) SIL
For 57 cells, we tested the effect of the presence or absence of internal contours. Comparison of the DLD and SIL conditions in Figure 6a and 6b shows that for two neurons removal of the dark occluding contours and of the inner lines separating the main parts of the objects had no effect on the neural responses and selectivity. At a population level, the neurons had similar response rates and selectivities under the two conditions. Figure 7c shows the distribution of RI for COL and SIL with a median of 0.19 (1\textsuperscript{st} quartile, -0.025; 3\textsuperscript{rd} quartile, 0.69; n=106), suggesting somewhat decreased firing rates for the SIL images. Twenty-one (20 %) of the neurons had RI over 0.8 suggesting sensitivity for the internal structure of the images. A comparison of the DLD and SIL responses revealed only very small differences (Figure 7d; median: 0.09 (1\textsuperscript{st} quartile, -0.02; 3\textsuperscript{rd} quartile, 0.31; n=76)).
4) LD

70% of the 44 tested neurons remained responsive to the LD stimuli when we removed all contrast from the images, but retained the contours as revealed by ANOVA (see Methods). However, this stimulus modification reduced the neuronal responses significantly. Typical neuronal responses are presented under the COL, DLD, SIL and LD conditions in Figure 6b. This response reduction was a general finding, as revealed by analysis of the 31 COL- and LD-responsive shape selective neurons. The median RI for the COL-LD comparison was 0.77 (Figure 7e; 1st quartile, 0.29; 3rd quartile, 1.17; n=44), suggesting significantly larger
responses under the COL than under the LD conditions (Wilcoxon matched pair test, $T=78$; $p<0.05$).

**ii) Effects of colour, texture, shading and internal contour removal on shape selectivity**

The shape selectivities of the neurons were also similar under the COL and the BLD and DLD conditions (Figure 8a). The average net responses under both texture-removed conditions decreased significantly with increasing stimulus rank (for this analysis, ranking was performed according to the neuronal responses under the COL conditions), demonstrating similar shape selectivities with and without internal texture information. The net response–shape rank curves, however, are flatter under the BLD and DLD conditions than under the COL condition (Figure 8a). To determine whether this is merely a consequence of the lower response rates seen under the surface-reduced conditions or constitutes a genuine difference in shape selectivity between the COL and BLD/DLD conditions, we additionally calculated the average net normalised responses (Figure 8b) dividing the responses by the response in the Rank 1 of the COL condition. Normalisation eliminates the absolute differences in net responses. As shown in Figure 8b, normalisation reduced the difference between the COL and BLD/DLD conditions. However, the decrease in the normalised firing rate with increasing stimulus rank is still significantly less under both the BLD and DLD texture-removed conditions as compared to the COL condition (ANOVA, interaction of rendering condition and stimulus ranking: COL-BLD: $F(3,267)= 20.65$, $p<0.01$ and COL-DLD: $F(3,225)= 35.87$, $p<0.01$), indicating that, at a population level, colour, texture and shading removal affected the shape selectivity weakly.

The PSTHs of two neurons whose selectivities are similar and whose selectivities are different under the COL and DLD (and SIL) conditions are presented in Figure 6a and b, respectively.

To determine how general is our finding, that shape selectivity is similar after colour, texture and shading removal we grouped our neuronal sample according to their behaviour. We defined 4 groups of neurons: neurons maintaining exact ranking order (1-2-3-4); neurons whose Rank1 is the same in the COL and in the surface reduced conditions (1-x-x-x); neurons whose Rank2 in the COL condition became Rank1 in the surface reduced condition
(2-x-x-x) and finally neurons whose Rank3 or Rank4 of the COL condition became Rank1 in the surface reduced condition. As it can be seen from Table 1 22 to 33% of the recorded neurons had exactly the same shape preference order in the COL and in the surface reduced conditions. We emphasize here that both stimuli leading to responses under Rank1 and Rank2 conditions were selected for greater effectiveness, while Rank3 and Rank4 conditions were selected as examples for ineffectiveness. This explains why both the 1-x-x-x and 2-x-x-x cell categories are consistent with generalization across rendering conditions. Thus, when considered together, around 80% of the neurons had the shape being defined as Rank1 or Rank2 in the COL condition as Rank 1 or Rank2 in the surface reduced conditions as well.

These data suggest robust independence of the neuronal shape selectivity from the rendering condition. As Figure 6a shows, for one neuron reversal of the contrast sign affects neither the neural firing nor the shape selectivity.

1) BLD-DLD

Analysis of the 57 COL-responsive and selective neurons revealed that this was a general finding: there is no significant difference between BLD and DLD in the firing rate – stimulus rank function (see Figure 8, ANOVA, interaction of rendering conditions and stimulus ranking: F (3,168)=1.04; N.S.), suggesting that selectivity is similar for stimuli brighter or darker than the background pattern, i.e. when the sign of the contrast between the object and the background is reversed.

Table 1

<table>
<thead>
<tr>
<th>COL-BLD</th>
<th>COL-DLD</th>
<th>COL-SIL</th>
<th>COL-LD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2-3-4</td>
<td>10 (33.3%)</td>
<td>15 (26.3%)</td>
<td>15 (26.3%)</td>
</tr>
<tr>
<td>1-3-2-4</td>
<td>24 (42.1%)</td>
<td>22 (38.6%)</td>
<td>24 (42.1%)</td>
</tr>
<tr>
<td>2-4-3-2</td>
<td>12 (21.1%)</td>
<td>13 (22.8%)</td>
<td>14 (24.6%)</td>
</tr>
<tr>
<td>3-4-2-3</td>
<td>2 (3.5%)</td>
<td>7 (12.3%)</td>
<td>4 (7%)</td>
</tr>
</tbody>
</table>

Key: 1-2-3-4, exact match of shape selectivity in the COL and in the surface-reduced conditions; 1-3-2-4, rank 1 is identical in COL and the other conditions, but the order of 3 and 4 is altered; 2-4-3-2, rank of COL became rank 1 of the other conditions; 3-4-2-3, rank 3 or rank 4 of COL.

Figure 8 Effects of elimination of texture and shading on shape selectivity of inferior temporal neurons. A) Averaged responses, B) Normalized responses
For 44 cells (responsive and selective under the BLD condition), we analysed the similarity of the shape selectivity under the BLD and DLD conditions further. For these cells, we ranked the 4 objects according to their net responses under the BLD condition, then we calculated the average net firing rate separately for each unit in the BLD and DLD conditions as a function of stimulus rank. We found no significant differences in selectivity between the BLD and DLD conditions (ANOVA, interaction of rendering condition and stimulus rank; F(3,129)=1.95; N.S.), showing, that the shape selectivity of IT neurons is independent of the contrast polarity of the image.

2) DLD-SIL

It is obvious from a comparison of the DLD and SIL objects presented in Figure 6a and 6b that removal of the contour lines from the images does not have equal effects on the perception of a simple, one-part object, such as a circle or of a more complex object with several different components, e.g. a drum. It is possible that the apparent lack of any difference we obtained under the DLD and SIL conditions is due to averaging of the response differences for simple objects and objects composed of several parts. To test this hypothesis, we made a separate analysis for those cells whose preferred stimulus (determined as Rank 1) is composed of at least 5 parts (i.e. the object could be separated into at least 5 closed, convex components by the dark inner lines under the DLD condition; e.g. stimulus #20 in Figure 2). However, there was no significant difference between the selectivities of these cells (n=29) under the DLD and SIL conditions (ANOVA, interaction of DLD and SIL (F(3,84)=1.31; N.S.), suggesting that this lack of difference in shape selectivity does not depend on the number of object components. The response strengths for these 29 neurons were also similar under the DLD and SIL conditions: RI for DLD and SIL with a median of 0.11 (1st quartile, -0.02; 3rd quartile, 0.27; n=29; Figure 7d).

![Figure 9](image-url) Comparison of the responses given to the conditions DLD and SIL
3) Overall

Figure 6b shows the shape selectivities for the COL, DLD, SIL and LD conditions for a TE neuron. At a level of population, the average net normalised response decreases significantly less with increasing stimulus rank under the LD condition as compared with the COL condition (ANOVA, interaction of rendering condition and stimulus ranking (F(3,63)=10.13; p<0.001; n=22), indicating that the removal of texture, shading and contrast affected the shape selectivity.

Figure 9 shows that, at a population level, the IT neurons exhibit similar selectivities for images with and without internal contours (ANOVA, interaction of rendering condition (DLD and SIL) and stimulus ranking: F (3,165)=1.07; N.S.).

4) LD

To analyse further the effect of contrast removal on the shape selectivity for 20 neurons (responsive and selective under the DLD condition) we ranked the 4 stimuli according to their net responses under the DLD condition. Next, we calculated the average net normalised firing rate separately for each unit under the DLD and LD conditions as a function of stimulus rank. Figure 10 shows the result of this analysis. The curve relating the average net response to stimulus rank is significantly flatter (ANOVA, interaction of rendering condition and stimulus rank, F (3,57)=3.65; p<0.05) under the LD than under the DLD condition. However, the average response to the preferred object under the LD condition is about 3 times higher than that to the non-preferred object, implying that IT neurons can signal objects depicted as line drawings.

Analysis of the neuronal sample revealed, that in case of LD 22% of the neurons Rank 3 or Rank 4 of COL became Rank 1 in the surface reduced condition, showing different selectivity for coloured shapes and contrast removed line drawings, a conclusion similar to the one obtained when comparing shape selectivity in DLD and LD conditions.
5) LD-COL

When we compared the rank1 COL responses with the rank1 responses of another cell-data-set which was obtained by searching for the cells with not COL but LD stimuli, we did not find differences in the amplitude of the responses for rank1 COL and rank1 LD stimuli.

The following tables (Table 2ab) and Figure 11 are showing the result of the comparison, giving the basic data of the recorded populations and the result of the t-test comparing them.

<table>
<thead>
<tr>
<th></th>
<th>Valid N</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>LD (res_sel_ld):</td>
<td>75</td>
<td>16,09208</td>
<td>-1,75000</td>
<td>62,60000</td>
<td>13,44702</td>
</tr>
<tr>
<td>COL (res_sel_col):</td>
<td>168</td>
<td>17,28480</td>
<td>-8,00000</td>
<td>56,40000</td>
<td>11,39803</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>t-value</th>
<th>df</th>
<th>p</th>
<th>F-ratio Var</th>
<th>p Variances</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0,711900</td>
<td>241</td>
<td>0,477215</td>
<td>1,391851</td>
<td>0,083874</td>
</tr>
</tbody>
</table>

Table 2ab

iii) Response latencies

The median response latency under the COL and BLD conditions was 103 and 110 msec, a difference not statistically significant (Wilcoxon matched pair test; T=1217; N.S.; n=79). The median response latency under the BLD and DLD conditions was 114 and 110 msec, again, a not statistically significant difference (Wilcoxon matched pair test; T=583; N.S.; n=49). The difference in median response latency was not statistically significant for DLD and SIL conditions either (110 msec under the DLD and 112 msec under the SIL conditions; Wilcoxon matched pair test; T=534; N.S.; n=50).

The median of the distribution of the neuronal latencies for the LD conditions was 101 msec, a value not significantly different from the median latencies of the same neuronal population.
under the COL and DLD conditions (Wilcoxon matched pair test, T=38 and T=39, respectively; N.S.).

iv) Analysis of different response intervals

The basic information about shapes is present in the very early part of the neuronal responses (Rolls and Tovée, 1994; Kovács et al, 1995b). Further, as found by Sugase et al (1999), IT neurons convey global information about the category (faces, shapes) of the stimulus in the earliest phase of their responses, while fine information about the identity or facial expression of the stimuli is conveyed later in the response. These two sets of data encouraged us to conduct a separate analysis on our data set. First, we determined the response latency of each cell under the COL condition for the stimulus, leading to the largest response. Second, we determined two response windows, a 100 ms long window, that immediately followed the response onset (Early) and a 100 msec window, starting at the end of the early response window (Late). Next, spike counts were computed off-line, trialwise for each stimulus condition with the previously determined 100 msec bins. Net responses were calculated by trialwise subtraction of the neural activity during a fixation period of 100 msec just preceding the stimulus onset. For an analysis of shape responsivity we separately computed RI indices for both the Early and Late response windows. None of the differences in RIs, obtained for the COL-BLD, COL-DLD, COL-SIL and COL-DLD comparisons, were significant (t-test for dependent variables, N.S.) between the Early and Late response windows. This suggests that information about the stimulus condition is similarly present in the Early and Late windows of the response.

A similar test was performed to determine whether the shape selectivity of the neurons is different for the Early and Late phases of the responses. We determined a selectivity index (SI). For each cell and each stimulus condition, we subtracted the average net firing rate in response to the least preferred stimulus (i.e. Rank 4) from the average net response to the preferred stimulus (i.e. Rank 1) under the same stimulus conditions and divided this difference by the sum of the two responses.

None of the SIs differ in the Early and Late response windows, suggesting that the shape selectivity is similar for the Early and Late response components.
c) Discussion

Our results can be summarised as follows. 1) Shape selective IT neurons remain selective for objects without texture and shading information. The responsiveness of the neurons, however, is affected by removal of these surface attributes. 2) IT neurons respond highly similarly to stimuli with opposite signs of contrast. Selectivity for shapes is also preserved over contrast reversal of the images. 3) Deletion of the inner contours has only mild effects on the responses and selectivity of the IT neurons.

i) Texture and shading

Few data are available as concerns the question of how a change of texture alters the shape selectivity of the IT neurons. In this study, instead of merely changing the texture, we removed all texture elements from within the objects. This stimulus variation affected the shape sensitivity of the IT neurons only weakly, suggesting the relatively low importance of texture in IT stimulus selectivity. However, the response rate did decrease under the texture-removed conditions, suggesting some degree of interaction of texture and shading with shape. This is in agreement with the conclusion of Vogels et al (1999), who systematically tested the effect of the angle of illumination on the IT neural responses and found that, for approximately half of the neurons, the direction of the illumination (i.e. the variations of shading) changed the neural activity.

ii) Silhouettes

The recognition of objects in ‘contre jour’ situations, when they are illuminated by a strong light from behind, can easily go astray (an example is that of children’s shadow-theatres). This shows that the outer or occluding contours of the objects alone are not always sufficient for proper recognition. On the other hand, schematic line drawings containing the inner contours that distinguish the main parts of the objects are at least as effective for object recognition as grey-scale or coloured representations (Biederman and Ju, 1988) At a neuronal population level, we observed similarly decreased firing rates for the objects containing the inner contours (DLDs) and for the SILs as compared to the chromatic versions, a result supported by another study (Vogels, 1999b). This suggests that inner contours are not necessary for the selective response of these neurons.
The explanation of the different effects of the elimination of internal contours on the behavioural and neuronal responses demands further studies. However, this discrepancy can be related to the different effects of the stimulus position on the behavioural performance and neuronal selectivity (Vogels, 1999b): changes of stimulus position led to responses similar to those for objects shifted in position, while the categorisation performance was affected strongly. It is possible that other neurons (within the IT or in different cortical areas) are responsible for the worse recognition of images presented in different locations and of SIL images.

iii) Contrast

In the real world, the sign of the contrast across the occluding contours of objects varies significantly, depending on factors such as the changing illumination and texture properties of the background. Nonetheless, perception is largely invariant to contrast changes in the objects. Indeed, real-time object-naming performance, long-term priming and immediate image integration processes are unaffected by the polarity of the contours and inner surfaces of non-face images (Subramaniam and Biederman, 1997). The comparison of our line drawing stimuli having higher (BLD) or lower (DLD) luminance values than that of the background showed no differences in either neural response rate or shape selectivity. This indicates that the responses of IT neurons do not reflect the contrast sign of the stimuli, suggesting that the IT may play a role in the contrast-invariant recognition of objects. This result is apparently in conflict with another report. Ito et al (1994) measured how the reversal of luminance contrast between object and background alters the neural responses in the anterior IT. Using the stimulus reduction method of Tanaka et al (1991) in anaesthetised animals, they found that for 60% of the neurons contrast reversal reduces the responses by more than 50%. Furthermore, 57% of their 19 recorded cells also displayed significant changes in shape selectivity with contrast reversal. They concluded that the IT neurons carry information about contrast polarity. The apparent disagreement between our finding and that from the study by Ito et al. can be attributed to the fundamental differences in the experimental approaches. First, Ito et al (1994) changed the contrast polarity of the objects and the backgrounds as well (i.e. they presented bright objects on dark surfaces or vice versa), while we presented our stimuli on an identical medium-grey background, making comparison of the two results difficult. Second, Ito et al used the stimulus reduction paradigm.
of Tanaka et al (1991), starting with a 3D object and eliminating step by step cues such as colour, texture and object-parts, in this way determining the critical feature for the neurons. During this process Ito et al. intentionally excluded those neurons that had texture or colour as critical features and studied only a small subsample of neurons that had their optimal stimuli defined exclusively by shape. This means that the neuron populations in the two studies overlapped only partially. Finally, Ito et al (1994) used anaesthetised animals, while we used awake, fixating monkeys. Although our animals were not engaged actively in any shape discrimination task during the recording sessions, we made attempts to draw their attention to the stimuli (see Methods)$^2$, making the correlation of the perceptual and neuronal results in our study a plausible one.

iv) LD

Removal of all contrast from within the objects and generating line drawings resulted in significantly lower response rates and changed selectivity. This result is in accord with the results of Ito et al. (1994), who also found changed selectivity for line drawing stimuli as compared to objects with surface cues.

1) LD as hunting stimuli

From our data it became obvious that the neuronal responses given to the LD condition are smaller compared to those given to the COL condition, and that the selectivity of the neurons is reduced or changed. This finding was however disturbing, since a number of studies (i.e. Sigala and Logothetis 2002) are using LDs as main stimuli, and they report much larger responses for them and a steep selectivity. Reconsidering the issue gave us the idea to compare responses to LD of those neurons, which had as “hunting stimuli” LD’s with the COL responses of those neurons which were found by using COL stimuli.

Although we miss one control condition, namely we do not have the data from those cells which had LD as hunting stimulus and COL as a recorded condition, we can draw a conclusion. For those cells which have been found by using COL stimuli, the responses given to the LDs are much smaller than those given to the same forms depicted as coloured

---

$^2$ In fact, during the recording sessions we had the common experience that, whenever the stimulus set was changed (from our standard set of 20 COL objects to the test objects under the COL, BLD, DLD, LD and SIL conditions), the animals had several ‘aborted’ trials for a while, as if they were ‘surprised’ by the sudden change of stimuli, showing the involvement of active attentional processes.
pictures, and the selectivity curve, when ranked according to the COL condition, was much less steep, almost flat. However, if the hunting stimuli are the LDs themselves, then the response given to the preferred stimuli is of the same amplitude as it was for the COL ones in the other case. This suggests that by searching the cells using coloured stimuli (which are supposedly preferred by IT cells) we introduce a bias. In other words we find a subpopulation of IT cells preferring coloured complex objects, and to some extent miss a possible other subpopulation, namely that of those cells which are tuned to line-drawn representations. Eventually the data suggest the (co)existence of (at least) two subpopulations of IT neurons as described above. This hypothesis is also in line with current perception-theories, suggesting either surface-based or edge-based methods in visual processing. Our data show that both approaches have their neuronal basis in the area responsible for object detection: the IT. The surface-bound cells prefer the coloured stimuli and they keep preferring them when the surface cues are reduced, as long as there is a surface to detect, but failing to respond when only the edges are visible (LD), and there is an other population which respond to those edges as vigorously as the others do to the surface-information.

v) Effect of practice

To test the possible effect of extended practice on the shape selectivity of our neuronal sample we analysed the temporal distribution of the selectivity of the neurons, which were responsive and selective for the two conditions under consideration. To do this we divided the total length of the recording period (that was done in 56 days for monkey1 and in 70 days for monkey23) into four, 15 days long periods. We performed a 3-way ANOVA of the normalized firing rates, ranked according to COL (dependent variables) and recording period, with repeated measure design as independent variable. This analysis suggests similar selectivity curves for COL and surface reduced conditions in each recording periods. This is an evidence for the absence of a significant effect of practice in the response selectivity of the recorded sample.

The presented physiological results regarding the shape selectivity of the IT neurons fit well with the psychophysical data of human and monkey experiments: both behavioural

3. Please note, that the number of the days, doesn’t correspond to the days in the calendar, i.e. we only counted the days of successful recordings.
performance and neuronal shape selectivity were largely invariant to the elimination of colour, to the inversion of contrast sign and to a lesser degree to the elimination of texture and shading. These results agree with the hypothesis that the IT plays a significant role in the discrimination and recognition of degraded images of objects under the variety of conditions encountered in natural environments, independently of the cues present in the image.

5. POSSIBLE CATEGORY-RELATED FEATURES IN THE RESPONSES OF IT NEURONS

a) Introduction

Previous studies have indicated that IT neurones respond differentially to complex visual stimuli (Gross et al., 1972, Desimone et al., 1984, Richmond et al., 1987, Eskandar et al, 1992).

It has also been shown that there exists a “complexity” gradient in the IT, the more anterior regions respond to more complex features (Tanaka et al., 1991, Tanaka 1996).

Tanaka et al., (1991) suggested that cells are clustered in cellular columns (modules) in the IT, containing cells selective for moderately complex features. They proved this idea by gradually reducing the images in their experiments to the simplest configuration, the critical features, which could still drive the neighbouring units. They hypothesized that general classes of objects are represented not by the activity of a single cell, but by the activity across different IT modules. Detailed discrimination would require the detection of small differences in activity of the neurones within the modules. Thus, IT cells form ensembles, involving different modules depending on the visual stimulus. New stimuli require a new recruitment of modules, which gives an infinite variability for coding novel stimuli using a relative small number of modules.

On the other hand, Sigala and Logothetis (2002) suggested that stimulus features important for visual categorization for instance, are represented in the activity of single units (neurones) in the primate IT cortex. In their opinion, neurones in the IT signal the diagnostic features for categorization via their firing rate, thereby, being especially selective for them.

One of the open questions concerning coding in IT is: how are objects similar to each other coded in the neuronal responses? Op de Beeck et al. (2001) used parameterised stimuli to
show that IT neurones represent similar shapes in a metrically biased but ordinally faithful way.

The present study is based on the further analysis of the experiments detailed above. During the experiments we presented 20 non-parameterised images similar in size and luminance but having different complexity (Figure 2), either geometrical shapes (GI) or photos of real objects (RI). The monkeys performed a fixation task while neuronal activity was recorded from the IT. The responses were subjected to cluster analysis, factor analysis and multidimensional scaling and the images were analysed regarding the physical features. Our results suggest that a given neuronal population in the IT may code simple and complex stimuli in different ways and similarity between images is manifested by the clustering of the neuronal responses given to them.

b) Specific Methods

i) Subjects
Subjects were the same monkeys as above.

ii) Statistics
The data files for the analysis contained 20 variables (columns) of the net responses to the 20 images (in Hz) and 117 and 100 rows corresponding to cells from Monkey I and Monkey II, respectively. The data from the two animals were analysed separately. Differences in all statistical procedures were considered significant at a probability level p<0.05 where applicable. Our analysis involved the following tests.

1. Cluster analysis (Ward method)
Cluster analysis may be used to classify a set of variables into groups (clusters) on the basis of their similarity (i.e. correlation, variance) or distance. Hierarchical cluster methods handle each variable as a single cluster in the first step. In each step, the two most similar clusters (variables) are collapsed, and similarities are recomputed. This process can be illustrated by a hierarchical tree or dendogram. From the several methods available, we used the Ward method, which uses an analysis of variance approach to evaluate the distances between clusters. This method attempts to minimize the sum of squares of any two (hypothetical)
clusters that can be formed in each step. The method is regarded as a very efficient one; however, it tends to create clusters that are small in size. Several definitions for distances can be applied; in this study, the squared Euclidean distance definition was used.

2. Factor analysis
The goal of factor analysis is to identify underlying variables (factors) that explain the pattern of correlations within variables. It involves the assumptions that the data should have a bivariate normal distribution for each pair of variables, and that the observations should be independent. Followed by a test of normality, the factor analysis was performed with the principal components extraction method and varimax rotation. The number of factors was determined, with eigenvalues over 1.

3. Multidimensional scaling (MDS)
MDS is used to detect meaningful underlying dimensions that allow the researcher to explain observed similarities or dissimilarities (distances) between the investigated objects (stimuli). MDS attempts to arrange objects in space with a particular number of dimensions so as to reproduce the observed distances. As a result, we can "explain" the distances in terms of the underlying dimensions. The computation uses a function minimization algorithm that evaluates different configurations with the goal of maximizing the goodness-of-fit (or minimizing the "lack of fit"). The most common measure used to evaluate how well (or poorly) a particular configuration reproduces the observed distance matrix is the „S-stress” measure. The smaller the stress value, the better is the fit of the reproduced distance matrix to the observed distance matrix.

A common way to decide how many dimensions to use is to plot the stress value against the number of dimensions. A second criterion for deciding how many dimensions to interpret is the clarity of the final configuration. However, if the data points in the plot do not follow any pattern, and if the stress plot does not show any clear "elbow," then the data are most probably random "noise". In this study MDS methods were performed on the basis of Euclidean distance.

The above methods do not result in significance levels, but are all useful to describe relationships between variables. Since they differ in method, we consider, that, provided the outcomes are similar, together they might lend support to our conclusions.
4. Physical features of the images

Since our stimuli did not differ in the obvious features such as mean luminance or surface area (see above), we looked for “non-obvious” characteristics, still in the physical domain. The following parameters were assigned to each stimulus: the surface area (SA), the perimeter length (PL), the total length of all the lines on the perimeter and also inside (AL), the perimeter length over the surface area (PL/SA) and the total length of all lines over the surface area (AL/SA).

Since we used colour stimuli, chromatic features were also analysed and the amount of colour information in the images was characterized in two ways. In one method, the change in energy of hue, saturation and brightness was calculated on the stimuli surfaces in the following way. First, the stimuli were transformed to HSV (hue, saturation, value) images with commercially available software. Next, the differences in the HSV values for neighbouring pixels were calculated along the horizontal and vertical axes of the images. Finally, we took the sum of squares of these values. This procedure resulted in three values for each image, as an expression of how much variability there exists on the surface in the hue, saturation and brightness domains, respectively.

In the other method, we used a “colourfulness index” (Tamura and Tanaka, 2001) for every image. Every pixel in the RGB images was given three values, according to the intensity of the red, green and blue components. Then the following formula was used:

\[
\text{Colourfulness} = \frac{\sum \sqrt{(\text{red-mean})^2 + (\text{green-mean})^2 + (\text{blue-mean})^2}}{M}
\]

where mean was \((\text{red+green+blue})/3\) and \(M\) was the image size in pixels (720*540). This procedure resulted in a single value, a “colourfulness index” for every image.

5. Sparseness index, selectivity

To quantify the stimulus selectivity measure of the tested neurons we used the sparseness (SP) introduced by Rolls & Treves (1990). This is a measure of the proportion of effective stimuli based on the response to each of the 20 stimuli. The sparseness indicates the length of the trail of the distribution of the net firing rates for the different stimuli (Treves & Rolls, 1991). Low values indicate long tail of distribution with only a few stimuli with high response rates. SP for \(n\) stimuli is computed by using the following formula:

\[
SP = \frac{\sum_{i=1}^{n} (R_i/n)^2}{\sum_{i=1}^{n} (R_i^2/n)}
\]
Where $R_i$ is the response to the $i$th stimulus of a stimulus set containing $n$ stimuli. $SP$ may range up to 1.0 indicating the case when a neuron responds to all of the stimuli in the stimulus set. We have to note that when calculating $SP$ the negative (netto) responses were clipped to zero.

Also, the selectivity indices (SI) were calculated for each neuron, see General methods, Data analysis section.

### 6. Latency times

A Poisson spike train analysis was used to detect and measure the latency of significant changes in activity in the epoch following target onset. In this method the actual number of spikes is compared to the number of spikes predicted by the Poisson distribution given the mean discharge rate of the cell. If the distribution of spikes is non-random in the spike train being analyzed, the Poisson spike train analysis determines the times of significant changes in the spike train (Legendy and Saleman, 1985, Hanes et al., 1995). There might be several times of modulation in a single trial. For each trial, times of significant neuronal modulation were collected. In an optimal case, at the end of a block of trials one would have several “candidates” for the latency values. The time across the trials at which this change was most likely to occur was given by the mode of the “time candidates” and was taken as the latency value of the response to visual stimulation. In those cases where for some reason there was no latency „candidate”, the trial was not included.
c) Results
We present data on 217 cells, 117 from Monkey I and 100 cells from Monkey II.

i) General findings
The average firing rate was 8.15 spikes/s for Monkey I and 11.6 spikes/s for Monkey II, which differed significantly from the baseline. As concerns the firing rates in response to GI and RI, no differences were found. The values were 7.9 spikes/s (SD ±10.3) and 8.4 spikes/s (SD ±6.4.) for Monkey I and 11.9 spikes/s (SD ±12.5) and 11.4 Spikes/s (SD ± 10.4 SD) for Monkey II, respectively. Distribution of the responses is shown on Fig. 12. The response levels are low, since many neurons decreased the firing rate to the stimuli we used.

The mean of the latency values in Monkey I was 114 ms (SD ± 21), identical to that in Monkey II: 114 ms (SD ± 16), t-test, p=0.99. In neither of the monkeys did we observe significant differences in latency times between the responses given to GI and RI (t-test, p=0.93 and p=0.89 for Monkey I and Monkey II, respectively). The distribution of the latency times is shown on Fig 13.

ii) Results of cluster analysis
Table 1 and Figs. 14 a and b show the different cluster solutions for Monkey I and Monkey II, respectively. A common property of the clusters is that generally images #1-11 and images #12-20 are “close” to each other and form two main clusters. The divisions between the two clusters correspond to the GI and RI stimuli. In the case of Monkey I, the separation follows the stimulus set: only the GI and RI stimuli can be found in one of the two main clusters. Also, in the case of Monkey II, GI and RI images
Figure 14ab

A

B
are completely separated. Interestingly there are images, which form pairs in both monkeys: stimulus #13 and #18 (the cross and the torso-like figurine) are next to each other in the dendrogram, as well as stimuli #1 and #5 (the triangle and the square), stimuli #3 and #10 (the star and the building blocks), and stimuli #14 and #16 (the cactus and the chalice).

Table 3.

<table>
<thead>
<tr>
<th>Proximity</th>
<th>Method</th>
<th>Clusters containing number of images</th>
</tr>
</thead>
<tbody>
<tr>
<td>M I.</td>
<td>Sq. Euclidean</td>
<td>Ward (1, 5, 2, 9, 8, 11, 3, 10, 4, 7, 6) - (12, 15, 20, 19, 13, 18, 14, 16, 17)</td>
</tr>
<tr>
<td>M II.</td>
<td>Sq. Euclidean</td>
<td>Ward (1, 5, 11, 2, 6, 8, 7, 3, 10, 4, 9) - (12, 15, 19, 17, 14, 16, 20, 13, 18)</td>
</tr>
</tbody>
</table>

First part (150 ms) of the response

<table>
<thead>
<tr>
<th>Proximity</th>
<th>Method</th>
<th>Clusters containing number of images</th>
</tr>
</thead>
<tbody>
<tr>
<td>M I.</td>
<td>Sq. Euclidean</td>
<td>Ward (1,5,11,2,8,7,3,10,9,4,6) - (12,18,13,12,16,15,19,20,17)</td>
</tr>
<tr>
<td>M II.</td>
<td>Sq. Euclidean</td>
<td>Ward (1,5,11,2,7,6,8,3,16,10,4,9) - (12,15,19,20,17,13,18,14)</td>
</tr>
</tbody>
</table>

Rest (350 ms) of the response

<table>
<thead>
<tr>
<th>Proximity</th>
<th>Method</th>
<th>Clusters containing number of images</th>
</tr>
</thead>
<tbody>
<tr>
<td>M I.</td>
<td>Sq. Euclidean</td>
<td>Ward (1,5,8,4,3,9,18,2,11,10) - (6,7,15,13,19,20,14,16,12,17)</td>
</tr>
<tr>
<td>M II.</td>
<td>Sq. Euclidean</td>
<td>Ward (1,7,6,8,20,2,18,17,16,11,15,3,4,5,9) - (10,13,19,12,14)</td>
</tr>
</tbody>
</table>

To see which portion of the cellular responses carries the information for clustering, the responses were divided into two segments, the first containing the first 150 ms of the response, and the second the remainder (the stimulus exposure time was 500 ms). We then repeated the cluster analysis for the two segments separately in both monkeys. We found that, while the first part gave consistent results in both animals, similarly to the original observation, i.e. GI and RI formed separate clusters (Table 3.); the second part of the response did not display the same clustering. The clusters contained a mixture of GI and RI and the two image groups were no longer separated.
iii) Results of the factor analysis

Monkey I.
Factor analysis resulted in 5 factors with eigenvalues over 1, which explained 69.16 % of the total variance. The rotated component matrix shows that the following images „form” a new factor (Table 4.).

<table>
<thead>
<tr>
<th>Factors</th>
<th>Stimulus number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>15, 12, 13, 20, 17, 18, 19, 14</td>
</tr>
<tr>
<td>2</td>
<td>10, 11, 8, 1</td>
</tr>
<tr>
<td>3</td>
<td>4, 9, 16, 2, 5</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>6, 3</td>
</tr>
</tbody>
</table>

Images #1-11 and #12-20 generally did not mix, and GI and RI remained separated, except in factor 3, where image #16 (chalice) was among the GI stimuli.

Monkey II.
Factor analysis resulted in 6 factors with eigenvalues over 1, explaining 72.37 % of the total variance. According to the rotated component matrix the following images “form” a new factor (Table 5).

<table>
<thead>
<tr>
<th>Factors</th>
<th>Stimulus number</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12, 14, 13, 19, 18, 16, 17</td>
</tr>
<tr>
<td>2</td>
<td>3, 15, 10</td>
</tr>
<tr>
<td>3</td>
<td>8, 6, 2, 5</td>
</tr>
<tr>
<td>4</td>
<td>11, 1</td>
</tr>
<tr>
<td>5</td>
<td>4, 9</td>
</tr>
<tr>
<td>6</td>
<td>7, 20</td>
</tr>
</tbody>
</table>
In this animal, some “outliers” were found, stimulus #15 (the duck) was among the GI images, and the last factor was formed by two images belonging in different classes: stimulus #7 belongs in GI, while #20 (the statue) is a member of RI. Other than that, GI and RI were separated.

iv) Physical features

We grouped our stimuli according to the results of the previous analysis; thus, stimuli #1-11 (GI) and stimuli #12-20 (RI) were put into separate groups. The physical features of the two groups were then compared. The mean surface size (in pixels) was 54440 in GI and 59220 in RI. No significant difference (t-test, p=0.33) was found, which was not surprising, since the stimuli were created in this way.

On average, however, RI had longer perimeter lines than GI (805 vs. 633 pixels, t-test, p=0.02) and the total length of lines (perimeter + internal lines) was also greater in RI than in GI (1357 vs. 2265, p=0.002) (Fig. 15).

We did not find any difference in the “change in energy” in the hue, saturation or brightness domains over the surface between GI and RI (hue: 0.3728 and 0.3728, t-test, p=0.24; saturation: 1.3601 and 0.3728, t-test, p=0.07; brightness: 1.1221 and 0.7660, t-test, p=0.27, respectively).

Analysis of the images in the RGB format, however, revealed that GI and RI differed in the mean colourfulness index (see Methods). This index ranged from 0.27 to 4.41, and GI on average had a higher index (2.52) than RI (1.45). This difference was significant (t-test, p=0.04).
v) **Sparseness, selectivity index**

Sparseness (SP) and selectivity indices (SI) did not differ between the two animals (Table 4). Also, there were no differences between SP of GI and RI stimuli. Between SI, however, a small but significant difference was found, namely, the indices were larger for GI than for RI ($t$-test, $p=0.04$). The distribution of SI and SP is shown on Figs. 16 and 17, respectively.

Table 6.

<table>
<thead>
<tr>
<th></th>
<th>SP simple (GI)</th>
<th>SP complex (RI)</th>
<th>SI simple (GI) ± Std</th>
<th>SI complex (RI) ± Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monkey I. (n=117)</td>
<td>0.42 (0.09-0.86)</td>
<td>0.51 (0.11-0.97)</td>
<td>0.56 ±0.17</td>
<td>0.50 ±0.19</td>
</tr>
<tr>
<td>Monkey II. (n=100)</td>
<td>0.44 (0.09-0.87)</td>
<td>0.51 (0.11-0.92)</td>
<td>0.60 ±0.18</td>
<td>0.53 ±0.18</td>
</tr>
</tbody>
</table>

**Figure 16** Distribution of the selectivity Index for the two animals in the two stimulus-groups

**Figure 17** Distribution of the sparseness of GI and RI
d) Discussion

Similarities of the responses from 217 cells in the IT given to 20 images were analysed by factor analysis, by cluster analysis and by MDS. A common result of each method was that on the basis of an analysis of the net cellular responses given to our image set, the geometrical shapes could be distinguished from those containing photos of real-world objects.

Tanaka et al. (1991) reported that the complexity required to drive the cells in the IT increased as the recording sites were moved from the posterior portion to the anterior part of the IT. Their study suggested that there might be a “complexity gradient” along the IT where neurones requiring different stimulus complexities are distributed in different anatomical locations. Optimal or non-effective stimuli were selected on the basis of the firing rates, i.e. stimuli were considered effective if they triggered high discharge rates, while non-effective ones caused only moderate increases in the firing rate or inhibition. Here, we report that cells in the IT might code stimuli with different complexities in other ways than just the amplitude of the mean firing rate.

In the present experiments, awake, behaving monkeys were used, while the Tanaka study used anaesthetized monkeys. It has been reported previously that the ratio of the responsive cells and the mean response magnitude is higher in awake than in anaesthetized animals (Tanaka et al., 1991, Tamura and Tanaka 2001). In our study, however, the mean firing rate was rather low. One possible reason for this might be the low luminance level of the images (around 8 cd/m² in our study as compared with around 35 cd/m² in the others) and that many of our cells were inhibited (i.e. gave negative net responses) to some of the stimuli.

Further, instead of trying to find the optimum stimulus containing the critical feature for the neurone, we presented all of our stimuli to all of the recorded IT neurones. This procedure could in theory produce differences in firing rates and separate the images on the basis of the amplitude of the neuronal activity but this was not the case in our study.

It could be argued that the differences in size or mean luminance between GI and RI can explain our results. The net responses and the latencies of the responses to GI and RI stimuli were compared and no significant differences were found (paired t-test). We consider that the differences observed are due to factors other than size or luminance differences since the simple (“first-order”) physical features did not differ in the two stimulus groups. We
hypothesized the presence of other, non-obvious, higher-order differences between the images used.

An attempt was made to identify some of the factors, which could explain the clustering of the responses to our stimuli. It is known for example, that the available surface information and the amount of internal lines influences the discrimination and recognition of images (Biedermann, 1987, Humphrey et al., 1994.). Indeed, our images differed both in the length of the perimeter and also in the total length of the lines visible in the stimuli.

Since the surfaces in the two stimulus groups did not differ, the line density per unit surface area was higher in RI (0.04 vs. 0.025, t-test, p<0.023, Fig. 14 a). The number of black and white transitions on the figure surface (essentially, internal lines) has been found to modulate the cellular activity in IT neurones (Eskandar et al., 1992), and thus the line density might be another factor responsible for the differences observed. Obviously, if two images have the same area, the one with the longer perimeter will appear more complex (e.g. a circle or square vs. a cross or star). Since these parameters add to the complexity of the images, we suggest, that IT neurones code images in a different way if the images differ in these characteristics. On the other hand, these factors might serve to express the similarity (in the physical domain) of images, which do not differ in features that normally influence response levels.

IT cells are best driven with complex, colourful images. Since our two stimulus groups differed in colourfulness index (2.52 and 1.45 for GI and RI, respectively), it seems that colour might be another clustering factor. Although there is evidence that many IT cells are selective for colours (e.g. Komatsu et al., 1992, 1993), our earlier studies revealed that the removal of colour information from the stimuli did not affect the responses of the neurones at a population level (Part 3 above, Tompa et al., 2005, Kovács et al., 2003). This suggests that at the level of the IT colour and shape processing occurs via different channels, though, the implementation of these channels is not yet clear. With different coding strategies the same single cells can process both kinds of information. Shape and form can be coded in the firing rate as responsivity and selectivity for shape. Colour information, on the other hand, cannot always be detected in the simple firing rate (Part 3., Tompa et al., 2005). Our finding shows that colour either independently or in combination with other characteristics, plays a role in the clustering of cellular responses. Thus, it is likely that the coding of colours does not rely simply on the change in mean firing rate.
It has been reported, that the early phase of the neuronal responses in the IT contains most of the useful information (Tovee et al, 1993). Following this logic, we divided our responses into two segments and the cluster analysis was repeated for the two segments separately in both monkeys. We found, that the first part gave consistent results in both animals, i.e. the geometrical shapes and real-world images formed separate clusters. The second part did not display the same clustering, the clusters were mixed and the two image groups were no longer clearly separated. This finding supports the idea of Tovee and others that IT neurones carry significant information about the visual stimulus in the first 100-200 ms of their responses. On the other hand, Sugase et al. (1999) demonstrated that IT cells can code information needed for different levels of categorization in the early and late parts of the neuronal responses, respectively. Our results, namely the difference in information content of the early and late parts can be the result of similar coding strategy.

In both animals, there were stimulus pairs, which grouped consistently in the dendogram: the torso-like statue (#13) and the crucifix (#18). It is tempting to speculate that this grouping was a result of some implicit categorization mechanism, and that the images were situated close together in the dendogram because of some semantical or higher-order categorisation process. We think that this is unlikely since the animals were required to perform a simple fixation task and the reward was probably associated more with the successful fixation than with the stimulus content. Further experiments are needed to clarify this question. Our data indicate the importance of the physical features rather than that of the content. In Monkey I for instance stimuli #2 and #9 (the square-ellipse, and polygon-circle) clustered, which might indicate the importance of the outline (Kovács et al., 2003) or building elements (see also stimuli #3 and #10 in both monkeys). Other examples could also be demonstrated, such as the internal pattern (images #1 and #5 in both monkeys), or the similar colors in stimuli 14 and 16 on both monkeys. These examples might be a manifestation of the coding mechanism Tanaka et al. suggested, that individual neurones code similar features, and small changes in the responses of individual neurones in a neuronal ensemble might result in the coding of the whole stimulus. It may also lend support to the results of Vogels (1999), who showed that whole neurone populations code a particular category (tree vs. non-tree). It is plausible to suggest that exemplars belonging in the same category share similar physical features (colourfulness, internal pattern, building elements colours etc.).
The sparseness values in our work did not differ much from those obtained in other studies like Vogels (1999) or Rolls and Tovee (1995). Vogels (1999) provides a relationship between AP level of IT and the sparseness. Values obtained in our work suggest two things: 1. The recording positions in our animals might be very similar to the most posterior electrode penetrations in Vogels (1999) and 2., the lack of difference between the sparseness values between our two monkeys gives an additional proof of similar electrode locations in both cases.

We found differences in the selectivity indices in both animals between GI and RI images, i.e., selectivity was higher for the geometrical shapes than for the real world photos. It is possible that this difference is due to the recording locations. The GI images are simpler from several aspects than the real world photos, and it is known that neurons in the anterior IT have a higher selectivity for more complex images than neurons in the posterior IT (Tanaka et al., 1991, Tanaka, 1996), where our data were collected.

Our results suggest that there might be separate coding in the monkey inferotemporal cortex for the processing of simple and complex images at a single cell level. Single cells might code the complexity of images reliably, even if they do not differ in simple physical features. Since the results from the two monkeys were highly similar, there might be general rules that determine the responses of IT cells to stimuli that differ in complexity.
6. CONCLUSION

In our work we have shown that the invariances experienced in the IT are indeed extended to most of the stimulus-reduced conditions – a result which is in line with the new dynamic view of the cortex, where the “early” areas provide a fine-grained buffer for the representation, and the higher ones (like IT or TE) are responsible for the immediate pre-categorization representation, regardless of the individual differences in the actual presentation. It remains a question if these invariances are acquired through training or are hard-wired. We found however an interesting exception to that rule – namely the depiction by lines: the line-drawn contours. It seems according to our results, that these (edge-defined) representations are dealt with by a separate population of cells in the IT. This on its turn suggests that there is a separate machinery in the visual system to process surface-defined and edge-defined representations, and this separation is present still in the IT. The same holds true to the color processing: we showed that at least for a considerable population of cells in the anterior IT, color does not add more information to what is already processed by the cell. This finding challenges the putative role of the inferior temporal cortex as a “colour center” in the macaque brain.

In our further investigations however it became clear that the population response of the IT cells indeed carries more information about the stimuli than what is evident from the study of the spike trains. The information we found seems to be contained in the difference of the variances in the neuronal responses across the area. The information we extracted turned out to be about the complexity of the given stimuli. Thus although the individual responses do not allow any fine-grained representation of the stimulus (it is not even needed: the representation is supposed to be given by the “active blackboard” or “buffer” of the “early” areas), the population response contains more information than provided simply by the frequency code.

Our further efforts shall be aimed to address the question of the reduction of our stimuli. Responses in the physical absence of the contours (illusory contours) will help us to elucidate the question of the surface-edge processing in the IT. We will also challenge the phenomenon detected in the first (“colour”) study: namely the putative task dependence of the responses (see figure 3) of the IT neurones.
REFERENCES


Hikosaka, K. Responsiveness of neurons in the posterior inferotemporal cortex to visual patterns in the macaque monkey Behav. Brain Res. 89:275-283 (1997)


Humphrey, G.K., Goodale, MA., Jakobson LS., Servos P. (1994) The role of surface information in object recognition: studies of a visual form agnosic and normal subjects. Perception. 23(12):1457-81,


ACKNOWLEDGEMENT

I would like to thank all of those who helped me in completing this work. First of all I would like to express my gratitude to the Head of the Physiology Department, Professor György Benedek for his valuable help both on scientific and human sides. I would also like to thank the founders of our monkey lab, Dr Gyula Kovács and Dr Gyula Sáry again from both professional and human aspects. I am thankful to my colleagues, Dr Zoltán Chadaide, Károly Köteles, Dr Gábor Csifcsák, Dr Laura Lenti for their work and help. They always created an very inspiring environment to work and learn in the lab. I thank our TDK-students, Dr Edit Franko, Edit Garab, Bencze Laczó and Dániel Hoffer for their great work. Thanks for the technical and personal assistance to Gabriella Dórai, Peter Liszli, also János Kóródi and Kálmán Hermann. Thanks to Dr Attila Nagy for his motivating presence in the lab. I am grateful to Mátra Janáky and Tamás Gyetvai for surgical assistance and to Erika Vörös for taking the NMR and CT images. I would also like to express my appreciation for our collaborators abroad, Rufin Vogels, Irving Biedermann and Jozsef Fiser.