
The Influences of Occlusion on Macaque Inferior Temporal Neurons

DISSERTATION

zur Erlangung des Grades eines Doktors
der Naturwissenschaften
der Fakultät für Mathematik und Physik
der Eberhard-Karls-Universität zu Tübingen

vorgelegt von

Kristina Juliane Nielsen
aus Hamburg

2005

Tag der mündlichen Prüfung: 13. Dezember 2005

Dekan: Prof. Dr. Schmid

Erster Berichterstatter: Prof. Dr. Ruder, Prof. Dr. Logothetis

Zweiter Berichterstatter: Prof. Dr. Schick

Abstract

Occlusion is ubiquitous in nature. It is therefore important for an observer to be able to identify visual input despite partial occlusion. Humans as well as Rhesus monkeys seem in general to be very good at this task. Yet, it has been demonstrated for human observers that the recognition of a partially occluded image depends on which image parts are visible. Here, it is shown that the same effects can be observed for monkeys. Two Rhesus monkeys were trained to discriminate between natural scenes. When the scenes were presented behind randomly generated occluders, the occlusion of very specific image regions impaired the performance. The results indicate that for monkey observers, information from different image regions contributes in a very characteristic way to the identification of an image. Each monkey had a unique bias to rely only on certain image regions during the task.

In the second stage of the experiment, single cell recordings were performed in area TE in the monkey inferior temporal cortex. The responses of neurons to partially occluded stimuli were recorded. Occluders were constructed according to the behavioral results, taking into account how behavioral relevance, or diagnosticity, was distributed across an image. The occluded conditions were constructed such that either diagnostic or non-diagnostic image regions were visible. Consistent influences of occlusion were present in area TE. Next to a general reduction in response rate with occlusion, pronounced differences were obtained between diagnostic and non-diagnostic conditions: Firing rate as well as selectivity was higher in the diagnostic conditions. The results therefore show that different regions of a natural scene evoke responses from TE neurons depending on their behavioral relevance.

Together with the behavior of single neurons, the local field potential (LFP) was recorded in area TE. The LFP reflects the synaptic activity in a brain region and thus can be considered to be coupled to the inputs into this region. While effects of occlusion were generally present in the LFP, the effects of diagnosticity depended on the recording location. Moving from posterior to anterior locations in TE, diagnosticity exerted progressively more influence on the LFP. Contrasting with the LFP,

the behavior of single units was homogeneous across the whole recording region. The data are consistent with a model proposing that diagnosticity is first encoded in posterior TE, but not in earlier visual areas.

In conclusion, we were not only able to demonstrate that the diagnosticity of an image region is encoded in area TE, but also that this encoding is first achieved there. In agreement with previous studies, this strongly suggests that one function of area TE is to store parts-based representation for learned visual structure, which reflects the diagnosticity of the encoded parts.

Zusammenfassung

Es kommt in der Natur häufig vor, dass nur Teile eines Objekts zu sehen, andere Teile aber verdeckt sind. Menschen sind trotzdem in den meisten Fällen imstande, die nur teilweise sichtbaren Objekte zu erkennen. Allerdings hängt es davon ab, welcher Teil eines Objekts zu sehen ist, ob es erkannt werden kann. Auch Rhesusaffen können in vielen Fällen nur teilweise sichtbare Objekte erkennen. Diese Studie zeigt, dass es dabei wie für Menschen entscheidend ist, welche Objektteile sichtbar sind. Zwei Rhesusaffen wurden darauf trainiert, zwischen mehreren natürlichen Bildern zu unterscheiden. Diese Bilder wurden dann hinter Masken gezeigt, durch die zufällig ausgewählte Bereiche der Bilder zu sehen waren. Die Abdeckung bestimmter Bereiche führte dazu, dass die Affen die abgedeckten Bilder nicht mehr erkennen konnten. Im Gegensatz dazu gab es Bildbereiche, deren Abdeckung sich nicht auf das Verhalten auswirkte. Die Affen stützten ihre Erkennungsleistung also auf spezifische Bildbereiche. Diese Bildbereiche werden im Folgenden als diagnostische Bereiche bezeichnet. Die Bestimmung der diagnostischen Bereiche zeigte große Unterschiede darin, welche Informationen die beiden Affen zur Erkennung eines Bildes heranzogen. Menschen verwendeten in den meisten Fällen andere Bereiche zur Erkennung derselben Bilder. Die diagnostischen Bereiche, die aus Experimenten mit Versuchspersonen ermittelt wurden, konnten daher nicht dazu verwendet werden, das Verhalten der Affen vorherzusagen. Des Weiteren zeichneten sich die diagnostischen Bereiche der Affen nicht durch besondere physikalische Bildeigenschaften aus. Weder die Helligkeit eines Bildbereichs, noch die Präsenz von Ecken oder andere Bildparameter konnten vorhersagen, welche Bildbereiche für die Affen diagnostisch sein würden. Bildbereiche wurden also nur deshalb diagnostisch, weil die Affen bestimmte Präferenzen für bestimmte Informationen bzw. bestimmte Strategien zur Lösung der Aufgabe hatten. Unsere Studie zeigt, dass diese Präferenzen bzw. Strategien sehr verschieden für Affen und Menschen sind. Für visuelle Aufgaben wird häufig direkt vom Verhalten der Menschen auf das der Affen geschlossen. Unsere Daten legen nahe, dass dieser Schluss nicht immer zulässig ist.

Von den Verhaltensstudien ausgehend wurden im zweiten Teil des Projekts Einzelzellableitungen im Areal TE im inferotemporalen Kortex der Rhesusaffen durchgeführt. Das Areal TE ist die höchste rein visuelle Verarbeitungsstufe und liegt im ventralen Temporallappen des Affenhirns. Es ist bekannt, dass Neurone in diesem Hirnbereich bevorzugt auf komplexe Objekte antworten. In dieser Studie wurde nun untersucht, wie sich unterschiedliche Abdeckung von Bildern auf das Antwortverhalten von Neuronen in TE auswirkt. Basierend auf den Verhaltensdaten wurden verschiedene Masken generiert. Masken wurden anhand von zwei Parametern konstruiert, der Diagnostizität und der Maskengröße. Die Diagnostizität bestimmte die Platzierung der Masken: Die Masken konnten so positioniert werden, dass trotz Abdeckung die diagnostischen Bereiche sichtbar blieben (diagnostische Bedingungen), oder aber diese Bildbereiche abgedeckt und andere Bildbereiche sichtbar wurden (nicht-diagnostische Bedingungen). Die Maskengröße dagegen bestimmte, wieviel vom Bild durch die Maske abgedeckt wurde. Der sichtbare Bildbereich betrug entweder 10, 30, oder 50%. Beide Faktoren hatten sehr unterschiedliche Auswirkungen auf das Verhalten der Affen. Die Diagnostizität der sichtbaren Bildbereiche korrelierte positiv mit der Wahrnehmungsleistung. Die Maskengröße dagegen beeinflusste die Wahrnehmungsleistung nur wenig. In diagnostischen Bedingungen war sie unverändert gut, auch wenn nur 10% des Bildes zu sehen waren. Nicht-diagnostische Versionen von Bildern wurden dagegen nicht erkannt, auch wenn die Hälfte eines Bildes zu sehen war.

Die teilweise Verdeckung von natürlichen Bildern hatte systematische Auswirkungen auf das Verhalten von Neuronen in TE. Die größten Unterschiede ergaben sich dabei zwischen diagnostischen und nicht-diagnostischen Bedingungen. TE-Neurone zeigten eine höhere Aktivität in den diagnostischen Bedingungen als in den nicht-diagnostischen Bedingungen. Gleichzeitig waren die Antworten der Neurone in den diagnostischen Bedingungen selektiver. Unterschiedliche Maskengrößen beeinflussten das Verhalten der Neurone ebenfalls systematisch. Diese Effekte waren aber schwächer als die Auswirkungen der Diagnostizität. Die Abdeckung eines Bildes hat also ähnliche Konsequenzen für das Verhalten der Neurone in TE wie für die Wahrnehmungsleistung des Affen. Ein Bild, das trotz Abdeckung zu erkennen war, führte auch zu einer Antwort im Areal TE. Sobald allerdings ein Bild aufgrund der Abdeckung nicht zu erkennen war, ging die Aktivität der TE-Neurone zurück.

Verglichen mit der neuronalen Antwort auf Bilder ohne Abdeckung war die Aktivität der Neurone für nur teilweise sichtbare Bilder generell reduziert. Die Aktivität wurde allerdings in diagnostischen Bedingungen weniger durch die Stimulusabdeckung vermindert als die Aktivität in nicht-diagnostischen Bedingungen. Insgesamt lässt sich also aus unseren Daten schließen, dass verschiedene Bildbereiche unterschiedlich

stark zur neuronalen Antwort auf ein Bild beitragen. Dieses Ergebnis stimmt mit früheren Studien überein, die ebenfalls zeigen konnten, dass unterschiedliche Bereiche von Objekten unterschiedlich stark zur Antwort von Neuronen in TE beitragen. Unsere Studie erweitert diese früheren Untersuchungen. Es war bis jetzt unklar, warum bestimmte Objektbereiche mehr Bedeutung für die Antworten in TE haben. Unsere Daten zeigen, dass die Diagnostizität eines Bildbereichs ausschlaggebend dafür ist, welche Aktivität der Bildbereich in TE hervorruft. Dieser Einfluss der Diagnostizität ist ebenfalls in Übereinstimmung mit früheren Studien: TE-Neurone zeigen z.B. eine höhere Selektivität für diagnostische Stimulusbereiche.

Gleichzeitig mit der Aktivität einzelner Neurone wurde an jeder Elektrode auch das lokale Feldpotential (local field potential, LFP) aufgezeichnet. Das LFP reflektiert die synchronisierten synaptischen Aktivitäten in der Umgebung der Elektrodenspitze. Die Analyse des LFPs beschränkte sich zunächst auf die Signalkomponenten mit einem festen zeitlichen Bezug zum Stimulus. Diese können durch Berechnung der sog. visuell evozierten Potentiale (VEPs) analysiert werden, die eine stimulus-bezogene Mittelung wiederholter LFP-Messungen darstellen. Drei VEP-Komponenten wurden genauer untersucht. Diese Komponenten wurden als N100, P130 und N200 bezeichnet, wobei die Notation die Polarität einer Komponente und ihre Latenzzeit nach Erscheinen des Stimulus angibt. Die teilweise Abdeckung eines Bildes wirkte sich auf alle drei Komponenten aus. In beiden Affen fanden sich Ableitungsorte, an denen die VEP-Amplituden der drei Komponenten rein zwischen diagnostischen und nicht-diagnostischen Bedingungen unterschieden. An diesen Orten hatte die Maskengröße keine Auswirkungen auf die VEP-Amplituden. Genauso gab es Ableitungsorte, an denen die VEP-Amplituden hauptsächlich durch die Maskengröße bestimmt wurde. Ableitungsorte, an denen sich ein starker Einfluss der Diagnostizität fand, waren hauptsächlich im anterioren Teil von TE konzentriert. Dagegen waren die Ableitungsorte, an denen sich ein starker Effekt der Maskengröße feststellen ließ, gleichmäßig über TE verteilt. Um diesen Effekt genauer zu quantifizieren, wurde für jeden Ableitungsort bestimmt, zu welchen Amplitudenmodulationen die beiden Faktoren im LFP führten. Übereinstimmend mit den vorigen Beobachtungen stellte sich heraus, dass der Faktor Diagnostizität einen zunehmend größeren Einfluss auf das LFP ausübte, je weiter anterior das LFP aufgezeichnet wurde.

Dieser Effekt wurde nur im LFP beobachtet; die Neuronen im gesamten getesteten TE-Bereich verhielten sich homogen. Das LFP reflektiert sowohl die lokale Aktivität am Ableitungsort, wie auch die Aktivität von entfernteren Hirnarealen, die Signale zum Ableitungsort übertragen. Es wurde nur innerhalb des Areals TE abgeleitet. Außerdem traten keine Unterschiede im Verhalten der Einzelzellen in verschiedenen TE-Bereichen auf. Deshalb kann davon ausgegangen werden, dass der lokale

Anteil des LFPs im posterioren und anterioren TE gleich bleibt. Damit bleibt nur der Signalanteil, der von anderen Arealen generiert wird, um die Abhängigkeit des LFPs von der Ableitungsposition zu erklären. In der Tat sind die Verbindungen für das posteriore und anteriore TE verschieden. Der größte Unterschied besteht darin, dass Verbindungen vom Areal V4 nur das posteriore TE erreichen. Insgesamt wird also das LFP im posterioren TE mehr von der Aktivität früher visueller Areale bestimmt als das LFP im anterioren TE. Der geringe Einfluss der Diagnostizität auf das LFP im posterioren TE lässt sich damit so interpretieren, dass die visuellen Verarbeitungsstufen vor dem Areal TE nicht zwischen Bedingungen mit verschiedener Diagnostizität unterscheiden können. Die Einzelzellen im posterioren TE reagieren aber, wie oben beschrieben, auf die Diagnostizität. Unsere Daten zeigen also, dass das posteriore TE die erste Stufe in der visuellen Verarbeitung ist, auf der die Diagnostizität eines Stimulus kodiert wird. Dieses Model erklärt nicht nur die hier vorgestellten Daten, sondern es ist auch in Übereinstimmung mit den Auswirkungen von Läsionen im Areal TE. Diese führen zu Defiziten in allen Aufgaben, in denen die diagnostischen Bereiche eines Stimulus identifiziert werden müssen. Insgesamt erweitert diese Doktorarbeit unser Wissen über das Areal TE also nicht nur darum, wie unterschiedliche Bereiche natürlicher Szenen kodiert werden. Die Daten zeigen vielmehr, dass die Ermittlung von Diagnostizität eine der speziellen Funktionen des Areals TE ist.

Contents

1	Introduction	11
1.1	Perception of partially occluded objects	11
1.1.1	Influences of occlusion on the perception of human observers .	11
1.1.2	Perception of occluded stimuli by animals	17
1.2	Properties of area TE	22
1.2.1	Anatomy	22
1.2.2	Connections of area TE	28
1.2.3	Consequences of lesions in area TE	31
1.2.4	General response characteristics of TE neurons	33
1.2.5	Receptive field properties	34
1.2.6	Stimulus selectivity	34
1.2.7	Effects of occlusion on the responses of TE neurons	44
1.3	Aim of the thesis	47
2	Methods	50
2.1	Animal preparation	50
2.2	Monkey psychophysics	53
2.2.1	Stimuli	53
2.2.2	Experimental paradigms	56
2.2.3	Apparatus	61
2.3	Human psychophysics	63
2.3.1	Subjects	63
2.3.2	Stimuli	63
2.3.3	Experimental paradigm	64
2.3.4	Apparatus	65
2.4	Neurophysiology	66
2.4.1	Stimuli	66
2.4.2	Task	70
2.4.3	Setup and recording procedure	71

2.5	Data analysis	74
2.5.1	Analysis of Bubbles data	74
2.5.2	Construction of a model observer	75
2.5.3	Eye movements	77
2.5.4	Image characterization	78
2.5.5	Analysis of neurophysiology data	82
3	Results	88
3.1	Behavioral results	88
3.1.1	Effects of occlusion on the discrimination of simple geometric shapes	88
3.1.2	Distribution of information across natural scenes for monkey observers	98
3.1.3	Partial occlusion of natural scenes - comparison between monkey and human observers	106
3.1.4	Eye movements versus Bubbles	115
3.1.5	Characterization of diagnostic image regions	123
3.2	Physiology results	130
3.2.1	Effects of occluder placement on single TE neurons	130
3.2.2	Influence of occlusion on neural selectivity	144
3.2.3	Influence of recording location on the behavior of single neurons	150
3.2.4	Representation of occluded stimuli by the LFP	156
3.2.5	Influence of recording location on the LFP behavior	162
3.2.6	Relationship between LFP and SUA	171
4	Discussion	179
4.1	Behavioral results	179
4.1.1	Effects of occlusion on the performance of monkeys	179
4.1.2	Methodological issues on Bubbles	183
4.2	Single unit responses	187
4.3	LFP responses	193
4.3.1	Representation of occluded scenes by the LFP	193
4.3.2	Comparison between the monkey LFP and the human EEG .	196
4.4	Comparison of single unit and LFP responses	201
4.5	Further experiments	203
5	Conclusions	206
	Bibliography	209

Chapter 1

Introduction

1.1 Perception of partially occluded objects

Stimulus occlusion effects were the main interest of this Ph.D. project. In the first section, the effects of the occlusion of objects are described, first for human observers, then for animal observers. The two subsections list the perceptual consequences of occlusion for the different observers.

1.1.1 Influences of occlusion on the perception of human observers

For instance, when two hills are visible far away, the base of one extending in front of the other and partly concealing it, we conclude immediately that the hill that is hidden is the more remote of the two; for if this were not the case, the form of the object would be different from that of any other hill that ever was seen; not to mention the strange coincidence that the outline of this peculiar hill should happen to be exactly continued by the contour of the other one. It might be a possible explanation of the picture presented to the eye, but it certainly would be contrary to all experience. (Helmholtz, 1910, p. 283, part 3)

In our everyday experience, visual information about the objects in our environment is often incomplete: Because of their three-dimensional nature, objects occlude parts of neighboring objects, as well as parts of themselves. However, we rarely notice the resulting lack of visual information. Instead, the visible object parts are identified as related, and the missing object pieces are filled in by the visual system, so that

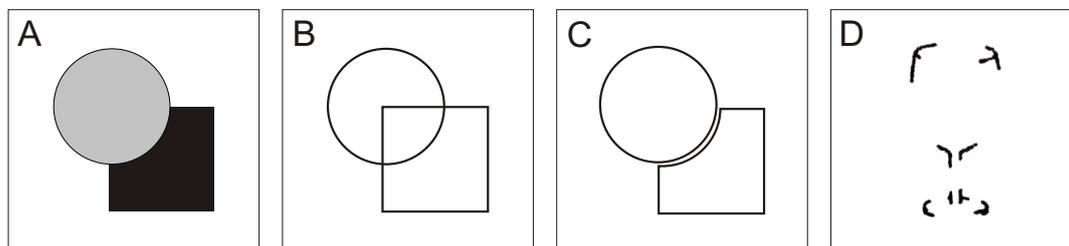


Figure 1.1: *Occlusion effects. A, Amodal completion. B - C illustrate two possible constructions of the stimulus shown in A by outlining their constituting elements. B, The typical complete interpretation. C, An interpretation of A as a two-dimensional mosaic. D, Perceptual closure. Even though stimulus parts are missing, the figure is completed into a wine glass.*

we perceive whole, uninterrupted objects. This process has been termed “amodal completion”. Incomplete information can also be generated by deleting parts of visual stimuli, as if placing them behind an invisible occluder. Again the missing parts do not prevent the perception of an uninterrupted object, a phenomenon termed “perceptual closure”. The perceptual impression generated in these cases is so realistic that when subjects were presented with complete and incomplete object pictures, they could not recall later which pictures had been shown as complete versions, and which ones hadn’t (Foley et al., 1997).

As an example for the effects of occlusion, consider Figure 1.1. Most observers will describe Figure 1.1A as a square lying behind a circle (Figure 1.1B), instead of a partial square adjoining a circle as in Figure 1.1C. Similar to the perceptual completion of the black region into a square, an observer automatically fills in the missing segments in Figure 1.1D and perceives a wine glass.

So far, most studies addressing occlusion phenomena have been concerned with amodal completion. The initial studies were restricted to demonstrations of the phenomenon, and were based purely on appeals to subjective evidence. However, it was already then noticed by Kanizsa that amodal completion possesses an automaticity that distinguishes it from stimulus completions constructed through logical reasoning of an observer. Kanizsa was also the first to demonstrate that amodal completion effects were accessible using quantitative measurements, so that the observation of the effect could be extended beyond purely subjective descriptions. Because of amodal completion, the size of a shape that is adjacent to a possibly occluding surface appears larger than the actual size of the shape (see Figure 1.2). In the experiment of Kanizsa, size judgements were given by observers, and a con-

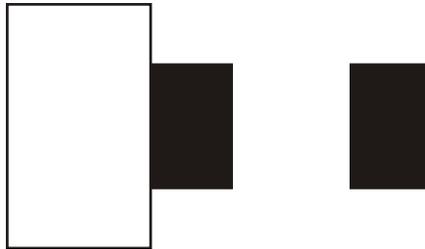


Figure 1.2: *The two black rectangles are equal in size; however, the one adjacent the white surface seems larger because of amodal completion.*

sistent expansion effect of about 8% could indeed be shown for occluded surfaces (Kanizsa & Gerbino, 1982). This provided a quantitative confirmation that amodal completion interacted with, and modified the perception of the physically presented information.

Since then, more quantitative studies have been performed, and amodal completion has been shown to greatly influence performance in many perceptual tasks, including motion perception (Shimojo & Nakayama, 1990; Joseph & Nakayama, 1999), visual search (He & Nakayama, 1992; Rauschenberger & Yantis, 2001), discrimination tasks (Sekuler & Palmer, 1992; Sekuler et al., 1994; Shore & Enns, 1997), and judgements about shape parameters (Murray et al., 2001). In these tasks, observers typically behave as if responding to the amodal completions of the stimuli, and not to the visible, unoccluded fragments. This happens even if using the unoccluded fragments would be beneficial for the task (He & Nakayama, 1992). Even though amodal completion proceeds automatically, it is not an instantaneous process. Using a priming paradigm, Sekuler & Palmer (1992) were first to show that completion requires a measurable amount of time. Further experiments have confirmed these results, yielding completion times from 80 to 200 ms, depending on the size of the occluded region (Shore & Enns, 1997; Murray et al., 2001). Before completion is reached, the low-level, partial representation seems to be available to the visual system, to be then replaced by the complete interpretation (Rauschenberger & Yantis, 2001, but see Bruno et al., 1997).

With respect to development, the perception of occluded stimuli seems to be an acquired process. Newborn infants are unable to perceive amodal completion. At the age of 4 months, this ability emerges first: When habituated to a rod moving behind an occluder, infants of this age looked longer at a subsequent display of two independently moving rod pieces than at a single, longer rod (Kellman & Spelke,

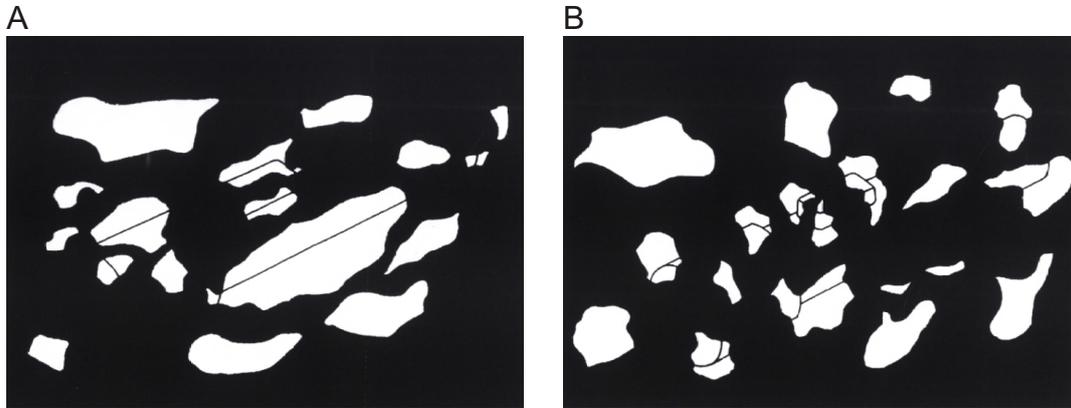


Figure 1.3: *Effects of occluder placement. A and B contain the same object partially visible through an occluder, but only in B the object is easily recognizable as a flashlight (from Biederman (1987)).*

1983). This indicates that the presentation of the occluded rod led to a perceptual impression of a single, moving rod. For a static display, infants have to reach the age of 6.5 months to be able to perceive the separate pieces of an object behind an occluder as connected (Craton, 1996). Finally, at around 8 months they begin to make assumptions about the form of the hidden object part (Craton, 1996). In agreement with these effects, infants aged 4.5 months could detect when an occlusion event contained conflicting information: They looked longer at a display in which a ball disappeared behind an occluder followed by the appearance of a box when the occluder was too small to cover both objects at the same time. At the age of 4.5 months, these object identity violations were detected when they involved shape; violations involving texture were noticed at the age of 7.5 months, and violations based on color with 11.5 months (Wilcox, 1999).

Even though the recovery of occluded objects seems to be easily possible under most circumstances, occlusion of different object parts can have different effects. Using line drawings of objects, Biederman (1987) deleted different parts of the line segments. Deletion of the middle section of a line segment was found to impair recognition performance much less than the deletion of line vertices. Using these results, Biederman constructed an example how for amodal completion the placement of the occluder could generate very different results. As can be seen in Figure 1.3, object identification depends on which part of an object is occluded.

More recently, Gosselin & Schyns (2001) systematically studied the effect of occluder placement on recognition performance. In their so called “Bubbles” paradigm, ob-

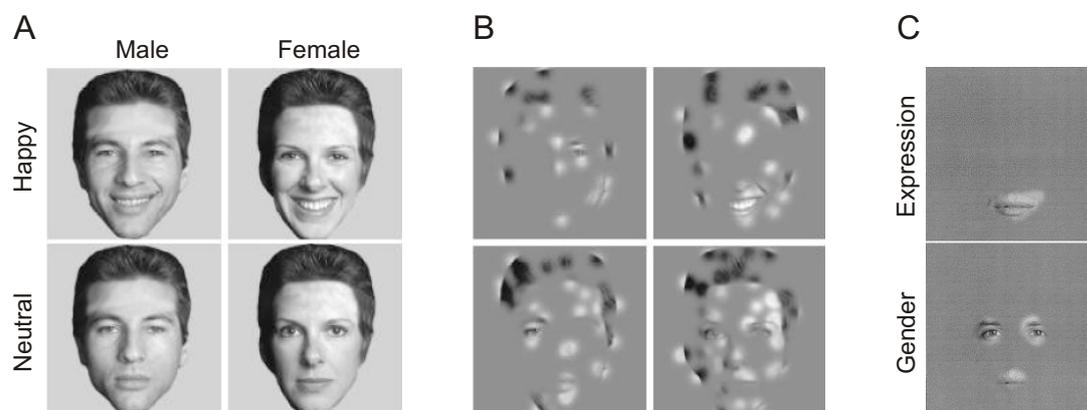


Figure 1.4: *A*, Exemplar face stimuli as used by Gosselin & Schyns. During the experiment, stimuli were shown behind masks with randomly placed windows, as illustrated in *B*. Occlusion of the face regions visible in *C* led to a disruption of performance in the expression discrimination (upper image) or the gender discrimination (lower image; adapted from Gosselin & Schyns (2001) and Gibson et al. (in press)).

servers were presented with a set of female and male faces, showing either a neutral or a happy expression (see Figure 1.4). Observers had to classify the faces either by their gender or by their expression. On every trial, the stimulus was shown behind an occluding surface containing round windows. The placement of the windows was random on every trial, and their number was adjusted to the performance of an observer. The results showed not only that the observers responded incorrectly in some of the trials, but also that the masks systematically interacted with the observers' performance. This interaction was furthermore dependent on how subjects had to classify the face images. While occlusion of the mouth region induced errors when faces had to be discriminated by their expression, subjects used eyes and mouth to determine the gender of a face. Occlusion can therefore prevent the identification of a stimulus, if the stimulus regions are occluded that subjects rely upon to perform a task.

Through completion processes, partially occluded stimuli are treated behaviorally as if they were completely visible. This incongruence between physically presented information and perception raises the question of how partially occluded stimuli are represented in the brain. Early experiments on this matter were adaptation experiments performed by Weisstein and colleagues (Weisstein et al., 1972a,b). In these studies, observers were first presented with different adaptation stimuli and subsequently had to judge the apparent contrast of a grating presented centrally. Subjects adapted either to a grating with a gap in the center, or to a grating with

a visible occluder at the gap position. In the latter case, the grating perceptually continued behind the occluder. Although the stimuli were physically almost identical, the amodally completed adaptation stimulus led to slightly larger adaptation effects. Adaptation effects were only observed if the adapting grating had the same orientation as the test grating. This experiment suggested that although the grating was only perceptually present at the center of the screen, it was nonetheless sufficient to adapt orientation sensitive channels in the brain, hinting at a neural representation of the amodally completed grating part.

The aftereffects generated by partially occluded stimuli are also consistent with the notion of a neural representation of the whole object instead of a representation of isolated parts: After adapting to a green rectangle partially occluded by two smaller rectangles, observers reported an afterimage consisting of a large red rectangle alternating with the afterimages of the smaller rectangles (Kanizsa & Gerbino, 1982). An afterimage generated by the physically presented information alone should have shown multiple, small red rectangles, in addition to the afterimages of the occluding rectangles.

The usage of functional magnetic resonance imaging (fMRI), as well as the recording of the electroencephalogram (EEG) allowed to study the brain activity for amodally completed stimuli in more detail. In an fMRI experiment by Lerner et al. (2002), higher order visual areas in the human brain were found to be only mildly affected by stimulus occlusion. The activation observed in the lateral occipital complex (LOC), a brain region responding to complex objects (Kanwisher et al., 1996; Kourtzi & Kanwisher, 2001), was reduced upon presentation of occluded object pictures in comparison to the activation evoked by the unoccluded pictures. However, the effect was smaller than the reduction observed after scrambling the objects to render them unidentifiable. It seems that as long as occluded objects are identifiable, the LOC responds similar to them as to the unoccluded images. In contrast, the activation of V4 seemed to be mainly determined by the physical properties of a stimulus and not whether it could be identified. For V4, adding an occluder increased the activation beyond the level observed for the unoccluded object; scrambling the object and thereby generating more discontinuities and corners increased the activation even further. Consistent results were obtained when the EEG was recorded while subjects viewed stimuli leading to amodal completion. The amodally completed stimuli generated an EEG scalp topography that was most likely generated from active brain regions in the LOC and the posterior parietal cortex. Again, no EEG sources were evident in the early visual cortex (Murray et al., 2004).

In the case of perceptual closure, the effects of missing stimulus information on object representation in the brain were studied with EEG recordings as well (Doniger

et al., 2000). In the experiment, observers were presented with line drawings of objects, of which parts had been deleted (the so called Snodgrass stimuli (Snodgrass & Feenan, 1990)). Upon first presentation of an image, large parts of an object were deleted, so that the objects were not identifiable. The presentation of the object was followed by less and less fragmented versions of the same picture, until just enough detail was presented to allow for object recognition. The EEG was recorded while observers viewed these stimulus sequences. The technique permitted the analysis of brain activity during object recognition, as well as during the completion processes preceding this stage. After the experiment, the visual evoked potentials were computed from the EEG data by averaging the EEG traces from repeated presentations of the same occlusion levels. The experiment showed that the magnitude of a negativity in the visual evoked potentials at about 290 ms after stimulus onset was influenced by the deletion of stimulus parts. This component could again be recorded above occipito-temporal areas. Adding physical information by presenting more of the stimulus parts increased the peak's amplitude gradually as long as stimuli could not be identified. Furthermore, when enough information was provided to allow stimulus identification, a larger increase in peak magnitude was observed.

1.1.2 Perception of occluded stimuli by animals

A number of animal species have been tested with respect to how they perceive partially occluded stimuli. Two outcomes are possible in this situation: Animals may only perceive the visible stimulus parts, or they may amodally complete the physically presented information, and perceive a larger object continuing behind the occluder.

Experiments have so far been performed on a number of species. They typically involved testing the animals on a discrimination task in which complete versions of a stimulus had to be distinguished from incomplete versions of the same stimulus. As an example, circles had to be discriminated from pacman stimuli. Partially occluded stimuli were inserted as test trials in the discrimination task, and it was recorded whether the animals categorized them as complete or incomplete. The results indicated that most animals tend to amodally complete partially occluded stimuli; however, there are some exceptions. Mice categorize partially occluded shapes as complete (Kanizsa et al., 1993). Chickens (Regolin & Vallortigara, 1995; Lea et al., 1996; Forkman, 1998) have been shown to amodally complete partially occluded stimuli; in contrast, an extensive number of experiments suggests that pigeons are unable to do so (Cerella, 1980; Sekuler et al., 1996; Fujita, 2001a). Chimpanzees (Fujita, 2001b) show behavior consistent with the perception of amodal

completion, but baboons only do so under certain circumstances (Deruelle et al., 2000).

Rhesus monkeys were tested in three different paradigms. In a discrimination paradigm, these monkeys have been shown to be able to discriminate shapes irrespective of whether they were shown in isolation or behind randomly placed masks (Kovács et al., 1995; Osada & Schiller, 1994), indicating that the masks did not disrupt the identification of the shapes. In another experiment, the study of Kanizsa described in Section 1.1.1 was repeated. It was confirmed that Rhesus monkeys systematically overestimate the length of a bar shown adjacent to a surface (Fujita, 2001a). Since the experiment allowed a parametrical assessment of the effects of occlusion on the perception of Rhesus monkeys, it provides strong evidence that Rhesus monkeys amodally complete partially occluded stimuli. Finally, Rhesus monkeys were trained to discriminate a long, continuous bar from two disconnected line segments by giving different responses to these two stimuli (Sugita, 1999). When the long bar was partially occluded by a surface, the monkeys indicated the perception of a continuous bar, consistent with the amodal completion of the two visible line segments.

The negative results for pigeons and baboons seem to stand out from the other experiments. However, these differences can possibly be explained as a failure to extract depth information from two-dimensional displays, and not as a failure to perceive amodal completion. The necessity of available depth information for amodal completion processes was suggested in a case study of Stevens (1983). The tested patient with a lesion in the left occipital and temporal cortex was unable to perceive amodal completion in a 2-D display. However, when stereoscopic depth information was added to the display, the patient's perception of partially occluded stimuli was normal.

In line with this hypothesis, chickens can identify which object is further away in a 2-D display depicting a spatial scene (Forkman, 1998), while pigeons seem to perceive all objects in a 2-D display in the same depth plane (Fujita, 2001a). Pigeons are also known to have problems with the extraction of depth information from 2-D line drawings (Blough, 1984). It might therefore be possible that pigeons could perceive amodal completion when tested with either the appropriate training or different displays. Support for this assumption is given by a study by DiPietro et al. (2002), in which pigeons were trained to identify four different objects. When after training the objects were shown upon or behind an occluding surface, pigeons failed in the identification task. Since this failure affected both presentation conditions similarly, it seemed to be generated by a difficulty segregating the object from the occluding surface, rather than a failure of amodal completion in the case of the occluded

presentation. Indeed, when the pigeons were provided with additional training on the objects now shown on top of an occluder during the training, they could transfer during testing to the same objects presented behind an occluder.

Similarly, baboons failed to associate partially occluded circles with their complete counterparts as long as stimuli were shown on computer displays (Deruelle et al., 2000, Experiment 3). However, when the presentation mode was changed, and stimuli were instead printed on cardboards which the baboons had to manipulate to receive rewards, occluded circles were associated with full circles (Deruelle et al., 2000, Experiment 4). Again, the availability of depth information is the distinguishing factor between both experiments. It therefore seems that – if the circumstances are appropriate – most animals perceive the visible parts of an occluded object as connected.

Only a single study has tested so far whether the occlusion of certain stimulus features selectively impairs stimulus identification. In a replication of the experiment by Gosselin & Schyns (2001), Gibson et al. (in press) used the Bubbles paradigm to test the ability of pigeons to discriminate partially occluded faces. As in the original study (see Section 1.1.1), pigeons had to discriminate human faces either by the faces' expression, or by their gender. Masks randomly revealed parts of the faces. Occluder placement had a systematic effect on performance, in that the occlusion of certain image regions affected stimulus discrimination more than occlusion of other regions. Pigeons, similar to humans, relied upon the eye region to identify faces by their expression. When faces had to be discriminated by their gender, pigeons tended to use both eye and mouth region, although these effects are not as pronounced as for human observers (see Figure 1.5).

The representation of partially occluded stimuli in the animal brain has been studied in Rhesus monkeys. Neural responses to partially occluded stimuli have been recorded in early visual areas V1 and V2, as well as in the inferotemporal cortex, a higher level visual area. The responses of neurons in the inferotemporal cortex to partially occluded objects are the topic of this thesis, and the general properties of this brain area will be described in the next section. In this section, only the results of the recordings in V1 and V2 will therefore be given, and the behavior of inferotemporal neurons when tested with partially occluded stimuli will be described later.

Sugita (1999) recorded from neurons in area V1. Neurons were selected which responded to a bar of a certain length. About 12% of these neurons stopped responding when the bar was replaced by two smaller, disconnected line segments, which if connected were as long as the original bar. Interestingly, the responses of these neurons

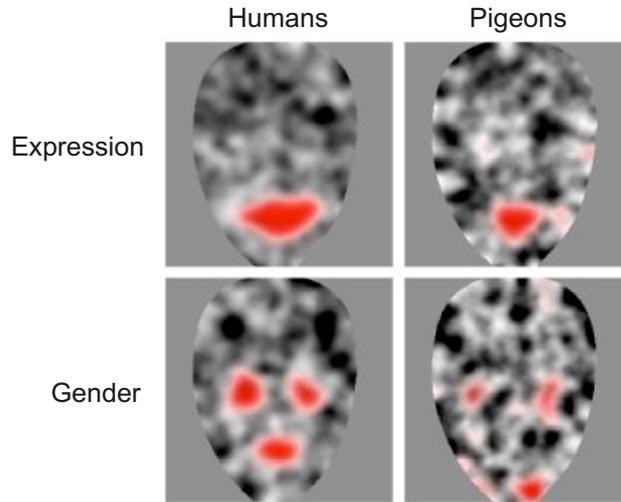


Figure 1.5: *Influences of occluder placement on the performance of humans and pigeons in a face discrimination task. Red regions indicate face regions in which occlusion disrupted performance. The upper row shows that humans and pigeons use the mouth region to discriminate faces by their expression, while the lower row contains the results for a gender discrimination (from Gibson et al. (in press)).*

could be restored if the same line segments were presented together with a larger surface which was placed in a depth plane in front of the line segments, and filled the space between them. Placing the same surface in a depth plane behind the two line segments left the neurons unresponsive. Perceptually, the line segments were perceived as one large bar because of amodal completion in the first condition, but not in the latter. The behavior of the neurons thus seemed to follow the perceptual impression, rather than the physically presented information. Bakin et al. (2000) confirmed these results in a more extensive study, in which neurons in V1 and V2 were tested. Many neurons in V2, but also a small proportion of neurons in V1 responded to illusory contours generated by occlusion. In conclusion, these results suggest that neural responses as early as V1 and V2 can show behavior sufficient to mediate amodal completion.

Finally, some conclusion about the responses of V1 neurons to amodally completed stimuli can be drawn from a study of Lee (2001) which tested different properties of V1 neurons. Nonetheless, from the example given in the paper (see Figure 3 of Lee, 2001), it can be concluded that neurons in V1 can respond to contours that are not physically present in a display, but generated because of amodal completion processes. The example also suggests that the onset of the responses to these

illusory contours is delayed with respect to the onset of responses to a real contour, suggesting that these responses are actually generated because of feedback input from higher visual areas.

1.2 Properties of area TE

So far, the general effects of stimulus occlusion have been described. This section gives an overview over the properties of area TE in the macaque monkey, to provide an introduction into the function of the brain area studied in this thesis. In the macaque monkey, information about visually presented objects is transmitted from primary visual cortex to inferotemporal cortex through multiple extrastriate areas. In this occipitotemporal pathway, the inferotemporal cortex (IT) is the final purely visual stage (Desimone & Gross, 1979). The first single-cell recordings in IT cortex were performed by Gross et al. (1969). Since this time, a number of neurophysiological studies have provided us with more detailed information about the properties of IT neurons. Reviews of these studies are for example given in Logothetis & Sheinberg (1996), and Gochin (1996). Here, after briefly reviewing the basic anatomy of TE, the effects of lesions of this brain area will be described, followed by a summary of the response properties of TE neurons.

1.2.1 Anatomy

IT is generally assumed to extend from the lower bank of the superior temporal sulcus (STS) to the lateral wall of the occipitotemporal sulcus (OTS). Horizontally, IT stretches from just anterior to the inferior occipital sulcus (IOS) to a couple of millimeters posterior to the temporal pole (Iwai & Mishkin, 1969; Desimone & Gross, 1979). Based on morphological data, neurons in this brain region are more complex than their counterparts in earlier visual areas: Layer III pyramidal cells in IT have both larger cell-bodies as well as larger basal dendritic fields as neurons in V1. While the somal cross-sectional area measures on average $100 \mu\text{m}^2$ in V1, the IT average is $230 \mu\text{m}^2$. The mean basal dendritic field area found in V1 is $25 \times 103 \mu\text{m}^2$, whereas its value in IT is $120 \times 103 \mu\text{m}^2$ (Elston & Rosa, 1998; Elston et al., 1999).

IT represents a large portion of the visual cortex, with a surface area corresponding to about 15% of the total area of the visual cortex (Felleman & Van Essen, 1991). Both anatomically and functionally, it is not a homogenous region. IT has consequently been subdivided, in most cases into the areas TEO posteriorly and TE anteriorly (Iwai & Mishkin, 1969; Kikuchi & Iwai, 1980; Boussaoud et al., 1991). A great variety of experiments, from lesion experiments, histological studies, to neurophysiological investigations, supports this distinction. These studies located TEO in a band of cortex running ventrally from the lip of the STS to about 2 to 8 mm medial to the OTS. The posterior border of TEO coincides roughly with the ante-

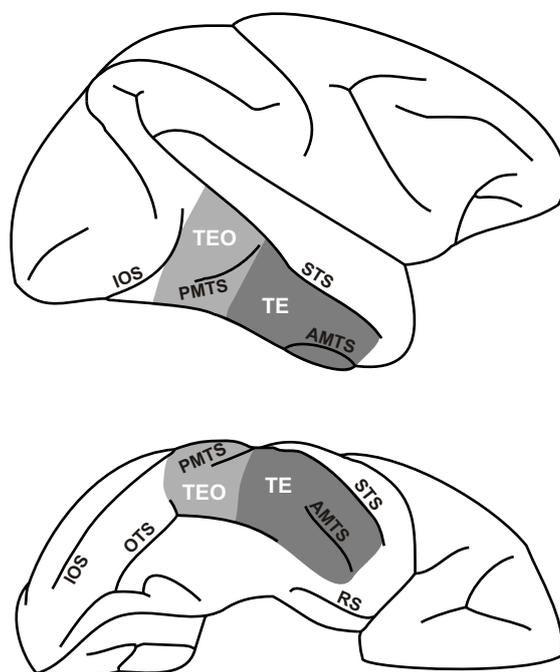


Figure 1.6: Locations of areas TEO and TE, sketched on lateral and ventral views of the right hemisphere of a typical rhesus monkey brain. Abbreviations: AMTS - anterior middle temporal sulcus, IOS - inferior occipital sulcus, OTS - occipitotemporal sulcus, PMTS - posterior middle temporal sulcus, RS - rhinal sulcus, STS - superior temporal sulcus.

rior lip of the ascending portion of the IOS. Anteriorly, TEO approximately ends at the anterior tip of the posterior middle temporal sulcus (PMTS) (Boussaoud et al., 1991). TE extends from the anterior TEO border, i.e. from the PTMS, to about the sphenoid, as shown in Figure 1.6.

Since regions within TE differ in their histology, and in their connections with other cortical and subcortical structures, a further subdivision of TE seems necessary. Using different criteria for the subdivision, two major partition systems for TE have evolved. The first scheme is illustrated in Figure 1.7. Here, TE is divided into four parts, namely TEpd (posterodorsal), TEpv (posteroventral), TEad (anterodorsal), and TEav (anteroventral). The demarcation between dorsal and ventral regions is placed at the anterior middle temporal sulcus (AMTS) and its caudal extension. The antero-posterior border is a vertical line crossing the posterior tip of the AMTS (Iwai & Yukie, 1987; Yukie et al., 1990). Subregions were identified as separate entities because the patterns of afferent and efferent connections vary across TE.

Specifically, dorsal and ventral regions are segregated by differing connections with the amygdala, the hippocampo-parahippocampal areas, and the perirhinal, prorrhinal and entorhinal cortex (Yukie et al., 1990; Saleem & Tanaka, 1996). Along the anterior-posterior axis, area TE is subdivided into two regions mainly by the presence or absence of direct afferents from area V4 (Shiwa, 1987; Morel & Bullier, 1990; Yukie et al., 1992).

The second partition system for TE was proposed by Seltzer & Pandya (1978). Based solely on the cyto- and myeloarchitecture within TE, they divided TE into five smaller regions, named TEa, TEm, TE1, TE2, and TE3. TE1 is the most rostral and ventral zone, situated immediately behind the temporal pole. Adjacent to it, progressively more caudally, lie TE2 and TE3. TEm straddles the lower edge of the STS, and TEa is located entirely within the lower bank of the STS. The partitioning scheme is depicted in Figure 1.8.

The two partition systems overlap partially: The border between TEav and TEad corresponds to the cytoarchitectonic border between TE1 and TE2. It is located at the lateral bank or lip of the AMTS at the rostrocaudal level, approaching the STS as it continues further anteriorly. Also, the lateral border of TEad is identical to the cytoarchitectural border between TE2 and TEm. Thus, TEav may be identified with TE1, and TEad with TE2 (Saleem & Tanaka, 1996).

The subdivision of IT into TEO and TE, with variations in the partition of TE, dominates in the recent literature. However, there exists at least one alternative partitioning scheme for the whole IT region, which was suggested by Felleman & Van Essen (1991). They divided IT into six subregions, forming three pairs of corresponding ventral and dorsal parts. Following their position along the anterior-posterior axis, the areas are labeled posterior (PITd, PITv), central (CITd, CITv) and anterior IT (AITd, AITv), as illustrated in Figure 1.9. The dorsal areas lie largely within the lower bank of the STS, extending a short distance onto the middle temporal gyrus. The ventral areas occupy most of the middle and inferior temporal gyri and extend into the lateral bank of the OTS. Borders between regions are delineated based primarily on changes in the connectivity of these regions. In addition, PIT subregions differ from CIT and AIT subregions in having a crude topographic organization.

Judging from their anterior-posterior level, PITd and PITv correspond to TEO, CITd and CITv to posterior TE, and AITd and AITv to anterior TE. However, the border between the dorsal and ventral parts of all three subregions is located close to the ventral lip of the STS. Since the lip of the STS represents the dorsal border of TEO as described by Boussaoud et al. (1991), PITv only corresponds to

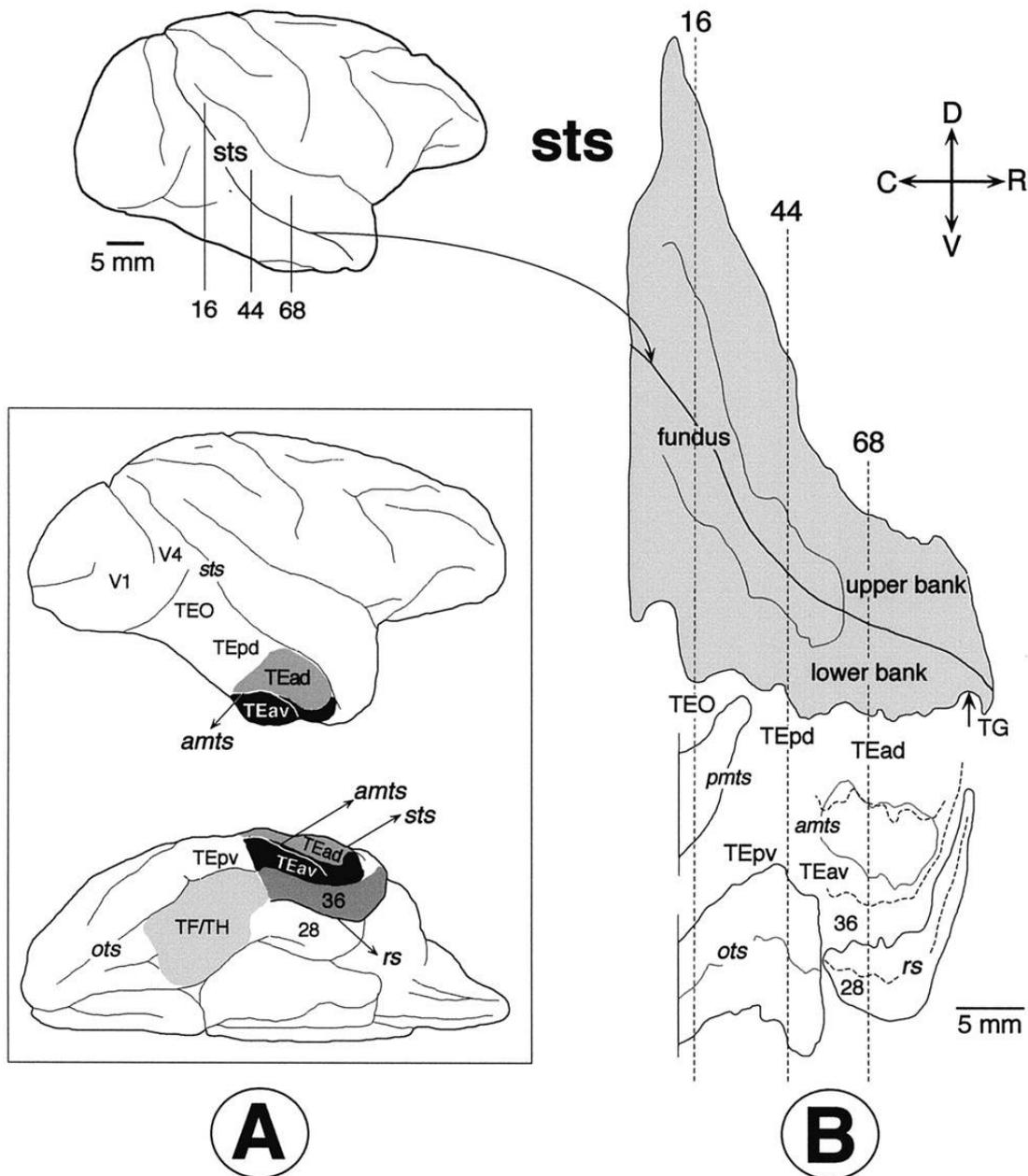


Figure 1.7: A, Lateral and ventral views of the right hemisphere of a macaque monkey, showing the locations of TEpd, TEpv, TEad, and TEav. B, Unfolded map, in which the STS, PMTS, AMTS, and RS are opened. Solid lines indicate lips and fundi of the sulci; broken lines show the borders between cortical areas (adapted from Saleem et al. (2000); abbreviations see Figure 1.6 and text).

TEO. PITd is often used referring to the ventral bank of the STS adjoining TEO. As another consequence, CITv overlaps both TEpd and TEpv, and AITv both TEad and TEav (Tamura & Tanaka, 2001). In comparison with the partition system of Seltzer & Pandya, CITd and AITd taken together incorporate TEa and TEm. CITv contains TE3 and parts of TE2, whereas AITv matches TE2 and TE1 (Felleman & Van Essen, 1991).

1.2.2 Connections of area TE

In most cases, the connections of either the whole TE region or of a single subregion in TE have been studied, prohibiting the analysis of differential projections originating in different subregions in TE. The results of these studies will be presented as connections for the whole TE region. Where applicable, differences between subregions will be specified.

TE receives its main input from area TEO, with additional input arising from area V4 and V4t (Desimone et al., 1980; Shiwa, 1987; Morel & Bullier, 1990; Yuki et al., 1992; Distler et al., 1993; Suzuki et al., 2000). The projections of V4 and V4t terminate in posterior TE only (Shiwa, 1987; Morel & Bullier, 1990; Yuki et al., 1992). Along the dorsal-ventral axis, visual afferents from TEO and V4 are differentiated roughly. Dorsal TE receives major projections from the representation of the central visual field in these areas, in contrast to a peripheral visual field input into ventral TE (Yuki et al., 1992). TE sends output back to V4, as well as to areas V1 and V2 (Suzuki et al., 2000). Of the areas in the caudal STS, TE has connections with FST (Morel & Bullier, 1990), PITd (Suzuki et al., 2000), PGa (Morel & Bullier, 1990) and IPa (Morel & Bullier, 1990). In the rostral STS, there exists a projection from area TPO in the upper bank of the rostral STS to anterior TE only (Morel & Bullier, 1990). In addition, the pattern of connections with the banks and fundus of rostral STS shows differences between TEav and TEad: TEad is interconnected with the upper bank of the rostral STS, as well as the rostro-caudally middle part, whereas TEav sends and receives information from the fundus and lower bank of the STS (Saleem et al., 2000).

TE also exchanges fibers with the temporal pole, which originate almost exclusively in anterior TE (Shiwa, 1987; Webster et al., 1991). Furthermore, parahippocampal regions TH and TF (Yuki et al., 1990; Webster et al., 1991; Saleem & Tanaka, 1996), as well as the perirhinal areas 36 and 35 (Yuki et al., 1990; Webster et al., 1991; Saleem & Tanaka, 1996), and entorhinal area 28 (Saleem & Tanaka, 1996) have reciprocal connections with TE. For each of these areas, the connection is stronger

with ventral than with dorsal TE (Yukie et al., 1990). From the inferior parietal lobule, TEav receives projections from areas PG, PF, and PFG (Zhong & Rockland, 2003). These are accompanied by reciprocal connections with areas LIPd and LIPv in the parietal cortex (Morel & Bullier, 1990; Webster et al., 1994). In the frontal cortex, area TE is interconnected with areas 8, 11, 12, 13 and 45a (Webster et al., 1994). Projections to the prefrontal cortex are maintained via the uncinate fascicle (Ungerleider et al., 1989). Within area TE itself, reciprocal connections have been shown between TEav and TEpv, as well as between TEad and TEpd (Yukie et al., 1992; Saleem et al., 2000). If the connections within TE are analyzed in the partition system of Seltzer & Pandya, reciprocal connections are found between TEa, TEm, and TE3. TE3 sends afferents to TE2, and TE2 to TE1 (Baylis et al., 1987).

The laminar pattern of the observed projections places TE at a higher hierarchical level than areas V1, V2, V4 and TEO (Morel & Bullier, 1990). FST and TPO are of the same level as TE (Morel & Bullier, 1990), whereas TG, TH, TF, and areas 35 and 36 follow TE in the hierarchy (Webster et al., 1991). The hierarchical relationship with IPa and PGa is unclear, but they have to be placed at least at the same level as TE, if not higher (Morel & Bullier, 1990). Also, the laminar pattern of projections does not allow an unequivocal determination of the relationship between TE and the prefrontal areas (Webster et al., 1994).

Besides its cortical connections, TE has a number of subcortical connections. TE receives input from the amygdala (Webster et al., 1991; Iwai & Yukie, 1987; Yukie et al., 1990; Cheng et al., 1997), the hippocampal formation (Webster et al., 1991; Yukie et al., 1990), and the lateral hypothalamus (Webster et al., 1993). However, in contrast to TEO, TE also sends fibers to the amygdala and the hippocampal formation (Webster et al., 1991; Yukie et al., 1990). The connections with the limbic system differ in their extent between dorsal and ventral TE regions. Dorsal TE is connected more strongly to the amygdala, whereas ventral TE is interconnected more with the hippocampal formation (Yukie et al., 1990). As area TEO, TE receives input from the locus coeruleus, the dorsal and median raphe, the basal nucleus of Meynert and the reticular formation (Webster et al., 1993). Additional projections originate in several thalamic nuclei, including the reticular nucleus, the nucleus reunions, and the pulvinar (Webster et al., 1993). Output from TE terminates in the striatum (Cheng et al., 1997; Webster et al., 1993) and the superior colliculus (Webster et al., 1993). Finally, TE also has a reciprocal connection with the claustrum (Webster et al., 1993).

Finally, TE makes connections with a number of contralateral structures, both cortical and subcortical. These concern the contralateral TE, amygdala, hypothalamus,

and midbrain structures (Webster et al., 1991, 1993). The interhemispheric connections are mainly maintained via the corpus callosum.

A number of predictions can be made based on the pattern of connections of TE and other cortical areas. First of all, TE receives its major inputs from V4 and TEO, but there are numerous other brain areas which provide input into TE. Consequently, neurons in TE should still be responsive even if the connections from V4 and TEO are severed. This has been confirmed by Buffalo et al. (2005), who reported normal firing rates in TE after lesions of both V4 and TEO. Second, with progression from V1 to V4 and TEO, neurons respond to more and more complex object features (Kobatake & Tanaka, 1994). Since based on the connectivity patterns TE has to be placed higher in the hierarchy than TEO, it can be expected that the properties of TE neurons are yet more complex. The next sections will prove this prediction to be correct.

Third, the functional properties of the areas that connect to TE should predict which TE behavior is mediated via these projections. Lesions of V4 and TEO have been linked to impairments in visual selective attention (De Weerd et al., 2003). The performance of monkeys with these lesions is normal as long as they are tested with stimuli presented in isolation; their deficits become apparent as soon as irrelevant stimuli are added to a display. Since for the lesioned animals information of relevant and irrelevant stimuli seems to be averaged together, it can be hypothesized that in normal animals, V4 and TEO filter this information out. With the lesion, information about the distractors reaches downstream area TE that normally would not have been transmitted. Indeed, the selectivity of TE neurons was found to be normal even after a V4 and TEO lesion as long as stimuli were presented in isolation (Buffalo et al., 2005). However, when distractors were added to the display, the selectivity of the TE neurons changed.

Similarly, conclusions about the role of the connections of TE to structures like the hippocampus can be drawn from the behavior of these brain regions. The hippocampal region, as well as the entorhinal, perirhinal, and parahippocampal cortices in the medial temporal lobe have been linked to the formation of declarative memory, i.e. the capability to recollect facts and events (for a review, see Squire et al., 2004). The functions of these brain areas are necessary to establish representations in long-term memory. However, they are most likely not the permanent repository of memory, which must be stored elsewhere. TE projects to and receives projections from these brain areas. Because of these projections, an involvement of TE in the storage of visual memory seems plausible. In this case, the backward signals from the medial temporal lobes should serve to consolidate the memory representations in TE. Indeed, Sakai & Miyashita (1991) demonstrated that after training

monkeys to associate pairs of geometric patterns, the responses of TE neurons to stimuli that formed a pair were correlated more strongly than to stimuli that were unrelated during training. These memory traces depended on the intactness of the feedback projections from the medial temporal lobe. When these connections were disrupted during learning, the associated stimuli did not evoke significantly correlated responses (Miyashita et al., 1996). In summary, the experiment confirmed that signals from the memory structures in the medial temporal lobe serve to establish long-term memory in TE.

1.2.3 Consequences of lesions in area TE

The behavioral changes of monkeys in which the inferotemporal cortex has been removed bilaterally implicate a participation of IT in the process of object recognition (for a review, see Dean, 1976, 1982). After IT removal, monkeys appear to have forgotten visual discriminations learned before the surgery, and they learn new discriminations much more slowly than normal animals. The deficit in discrimination learning is very long lasting, and represents the most pronounced effect of IT lesions. It affects a wide variety of discriminanda, including both very simple and complex patterns. The impairment is restricted to vision, with discrimination learning for other modalities being unaffected. In general, the severity of the deficit caused by an IT lesion increases with task difficulty. It is not attributable to a “simple sensory loss”, since IT lesions that impair visual discrimination learning do not reduce visual acuity or produce detectable scotoma. In addition, perceptual thresholds have been measured along a number of dimensions, and were found to be normal in each case. Most recently, thresholds have been shown to be normal for shape distortion when animals had to detect a distorted shape in a set of simultaneously presented shapes (Huxlin et al., 2000).

Insights into the sources underlying the observed deficits come from studies by Iwai and colleagues (Iwai, 1985; Iwai et al., 1990). In these experiments, monkeys were taught a discrimination task using a Wisconsin General Testing Apparatus. In the apparatus, patterns were displayed on large plaques that covered two food wells in front of the monkeys. The monkeys had to touch the edge of a plaque to push it aside, and to retrieve a bait possibly hidden under the plaque. On each trial, the same two stimulus cards were presented, one of which was consistently associated with a reward. When the monkey’s learning curves were analyzed under these conditions, they could be shown to consist of two separate stages. During the first stage (Learning Stage I), the monkey’s performance remained at chance level, while in the second stage (Learning Stage II) the number of correct responses increased

drastically. At the end of Learning Stage II, monkeys could discriminate the patterns without problems. Further experiments showed that the duration of Learning Stage I could for example be reduced if the monkeys were trained with one pair of patterns first, and subsequently learned to discriminate a second pattern pair. Learning Stage II however was unaffected by such manipulations. The authors concluded that Learning Stage I reflected the learning process to attend selectively to the discriminative cues shown on the plaques. Only in Learning Stage II, animals seemed to actively sample the patterns and to identify the discriminative features of each pattern. In normal, naïve animals, Learning Stage I accounted for most of the time necessary to learn the discrimination task, yet in animals with IT lesions, Learning Stage II was significantly longer than Learning Stage I. It was also longer than the Learning Stage II in normal animals, whereas the duration of Learning Stage I was not affected by IT removal. The results indicate that IT lesions do not affect attention mechanisms, but instead lead to a deficit in pattern perception, or the identification of the task-relevant pattern features.

A similar conclusion was drawn by Gaffan et al. (1986), who also tested monkeys with IT lesions on a discrimination task. The monkeys were taught to discriminate between pairs of stimuli. In each stimulus pair, one stimulus was selected as the “positive” stimulus. The two stimuli were presented at the same time; if the monkey selected the positive stimulus, reward was delivered. Each stimulus pair was presented for 100 trials before a new pair was introduced. In this way, the learning performance for a total of 60 stimulus pairs was obtained. Monkeys without a lesion were tested on the same stimulus pairs for comparison. While the general deficit in discrimination learning was confirmed for the lesioned animals, it was also observed that their performance varied widely between stimulus pairs. In many cases, there were no errors at all, in others, over 70 trials of the block of 100 trials were incorrect. A more precise analysis of the individual discriminations showed that the lesioned animals expressed very strong preferences for one or the other stimulus during the first few trials with a new stimulus pair. These preferences persisted throughout the whole block, even if the monkeys preferred the unrewarded stimulus. Therefore, performance for a particular discrimination problem depended on the initial preference of a monkey. If the monkey initially preferred the positive stimulus, error rates were generally low, while preference of the negative stimulus led to high error rates. No such preference was observed for the control animals. Interestingly, the lesioned animals agreed in their initial preferences. Furthermore, the degree of difficulty of a particular problem was not correlated between control and lesioned animals.

The data are consistent with the hypothesis that IT lesions reduce the set of features available to describe a shape. As a consequence of a reduced feature set, lesioned

animals will find other problems difficult than normal animals. In addition, if stimuli are described by few attributes only, inappropriate transfer between discriminations learned at different times becomes very likely because of the overlap of attributes between objects. In these cases, attributes of new stimuli carry over irrelevant associations from previous problems. A reduced feature set does therefore not only impair the learning of new discriminations, it may also be reflected in a strong preference for a particular shape in a completely new discrimination problem. The authors summarize their results in the conclusion that the IT cortex enables normal animals to take advantage of a wide range of attributes which in combination identify a particular complex visual input.

Most recently, IT lesions have been shown to affect performance in oddity tasks (Huxlin et al., 2000). In an oddity task, a number of stimuli are presented simultaneously. All stimuli except one are identical, and the monkey is required to locate the odd stimulus. Oddity tasks are an extension of stimulus detection tasks, in which the monkey has to locate a predefined target amongst a number of distractor stimuli. Postoperatively, animals could eventually be retrained to perform stimulus detection tasks. However, they did not relearn the oddity tasks within the duration of the study. The crucial difference between these two paradigms is the way in which the response target is defined. In stimulus detection tasks, the target is predefined and identical on every trial, while in oddity tasks, a target is only defined through its relationship with the other stimuli, and it changes on every trial. Oddity tasks therefore rely more heavily on the ability to identify the characteristic attributes of an object. Again, the results point towards an involvement of IT in the extraction of the task relevant shape features.

1.2.4 General response characteristics of TE neurons

Neurons in TE are exclusively visual (Gross et al., 1972; Baylis et al., 1987; Desimone & Gross, 1979). They respond to visual stimulation with an onset latency of 50 to 250 ms (Hikosaka, 1999; Baylis et al., 1987; Sheinberg & Logothetis, 2001; Tamura & Tanaka, 2001), with a peak latency in TE of about 150 ms (Tamura & Tanaka, 2001). Neurons in TE have been reported to have variable and bursty spontaneous activity (Desimone & Gross, 1979). In addition, the responses of TE neurons were found to decline with repeated stimulation if the time interval between the stimulations was shorter than about 5 s (Gross et al., 1972; Desimone & Gross, 1979). Most studies have analyzed changes in mean firing rate to characterize the behavior of IT neurons. However, the time course of the responses has been proposed to carry information as well (Richmond & Optican, 1987). Interestingly, neuronal selectivity

might be time dependent, with neurons becoming more selective over time (Tamura & Tanaka, 2001).

1.2.5 Receptive field properties

TE neurons were initially reported to have receptive fields as large as 60 deg (Gross et al., 1977; Desimone & Gross, 1979; Boussaoud et al., 1991; Kobatake & Tanaka, 1994). More recent studies have however shown that receptive field size in TE depends on the stimulus size (Op de Beeck & Vogels, 2000), and may be much smaller. Using stimuli with a size of about 3 deg, Op de Beeck & Vogels (2000) reported an average receptive field diameter of 10 deg. In their sample, the smallest receptive field had a diameter of about 3 deg, while the largest diameter equaled 26 deg. Even smaller receptive fields were observed by DiCarlo & Maunsell (2003), who measured position dependencies in TE consistent with receptive field diameters of 2.5 deg when testing with very small stimuli (0.6 deg width).

No retinotopic organization could be identified in area TE (Desimone & Gross, 1979; Boussaoud et al., 1991). The receptive fields found in TE almost always include the center of gaze (Gross et al., 1969, 1972; Desimone & Gross, 1979; Op de Beeck & Vogels, 2000), with neurons responding to stimulations both in the upper and lower visual field (Boussaoud et al., 1991). A large number of TE units has bilateral receptive fields (Gross et al., 1969, 1972, 1977; Desimone & Gross, 1979). These neurons receive information from the ipsilateral visual field by connections from the opposite striate cortex through the anterior commissure and the splenium (Gross et al., 1977).

1.2.6 Stimulus selectivity

The stimulus selectivity of IT neurons was initially tested by comparing the responses evoked by different classes of stimuli, such as bars and disks, or complex shapes (see for example Baylis et al., 1987). By this means, it could be established that most IT neurons do not respond to simple stimuli, and instead can only be activated by the presentation of complex shapes (Gross et al., 1972; Baylis et al., 1987; Desimone et al., 1984; Tanaka et al., 1991). Figure 1.10 shows recordings from one such IT neuron that maximally responds to stimuli with irregular edges, but cannot be activated by simple, more regular shapes. The preference for complex stimuli seems to be especially pronounced in area TE (Kobatake & Tanaka, 1994).

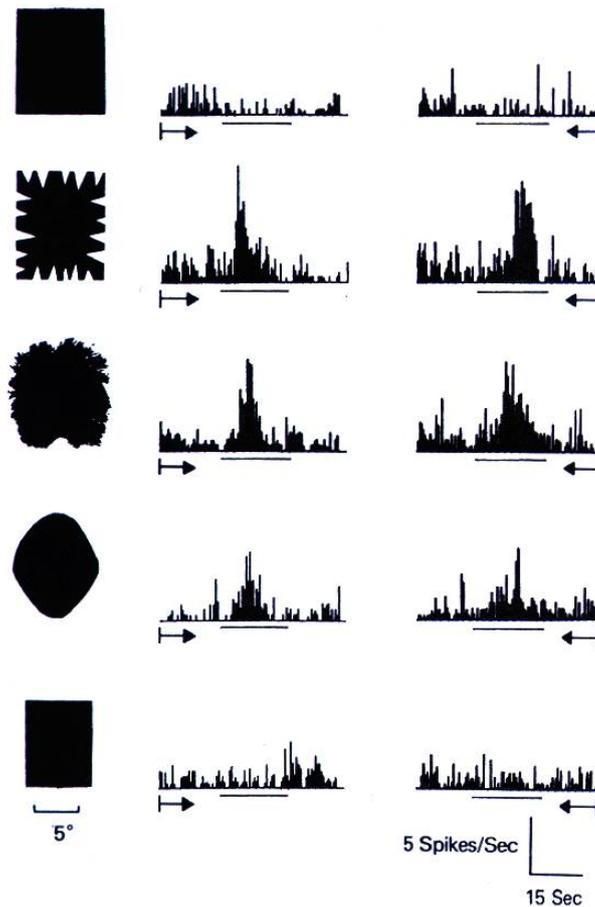


Figure 1.10: *Peristimulus time histograms (PSTHs) for an illustrative IT unit selective for patterns with irregular edges. The bar indicates the time during which the stimulus was shown, and arrows the direction of stimulus motion and the direction of time in the histograms (Desimone et al., 1984).*

This property clearly distinguishes TE from earlier visual areas. Its complex trigger features make TE a suitable candidate for processes involved in object recognition.

Upon presentation of a large number of stimuli, some TE neurons could be identified as selective for a single class of objects or object properties. As an example, so called face cells have been identified in TE, for which the sight of a face elicited two to ten times larger responses than other visual stimuli (Perrett et al., 1982). Face selective neurons responsive to the identity of faces were found in the inferior temporal cortex, while cells that responded to facial expressions, gaze direction, and vantage point were mostly located in the STS (Hasselmo et al., 1989; Perrett et al., 1991, 1992). This clustering of cells with the same properties in some TE parts has raised the question of whether TE possesses a particular functional architecture.

Only one study could so far demonstrate that neurons in different subregions in TE indeed differ in a specific functional property. In the experiment by Janssen et al. (2000), the responses of TE neurons to 2-D and 3-D shapes were compared. A clear distinction between neurons in the lower bank of the STS and neurons in more lateral TE parts became apparent: 56% of the neurons in the STS distinguished between shapes that were similar with respect to their 2-D projections, but differed in their depth information. In contrast, only 12% of the neurons in lateral TE were selective for the 3-D aspects of a shape.

For the majority of TE neurons, no preference for a specific stimulus class was observed. Most neurons responded to a broad range of objects, limiting the identification of further subdivisions within TE. A rather general distinction between areas TEav and TEad emerged from a study by Tamura & Tanaka (2001): Neurons in these two regions were tested with a set of 100 different stimuli, with maximally ten stimuli belonging to the same class. For the average neuron in TEad, 17 of the 100 stimuli evoked discharge rates significantly higher than the spontaneous activity, demonstrating a broad neuronal selectivity. The median of the distribution was at 7 stimuli evoking responses significantly larger than the baseline activity. In contrast, the median for the population of TEav neurons was 4 stimuli triggering significantly enhanced firing rates.

Because the observed stimulus selectivity in TE is so broad, it is difficult to identify which stimulus dimensions determine the neuronal responses. One approach to the problem was based on the assumption that responses to different objects are actually triggered by a part or feature common to these objects. To identify this feature, the responses to object parts need to be compared to the responses to the full objects containing these parts. This line of research was extensively followed in the lab of Tanaka (for a review, see Tanaka, 1996). In their experiments, 3-D objects were used to stimulate TE neurons. If a neuron was found to be responsive to a certain object, the object was photographed. The object's image was subsequently simplified to identify the object feature that the neuron responded to (the critical feature). Simplification steps included the removal of color information, texture, gradual luminosity change, and highlight, as well as the transformation of the photography into a simpler geometrical figure. In subsequent steps, the critical aspect of this figure was determined by modifying the shapes of components or the overall configuration, or by decomposing the image into components (Fujita et al., 1992). After each simplification step, the neuron's response to the simplified stimulus version was compared to the response to the original version. Testing proceeded only if the two responses were comparably strong, arguing that in this case the removed stimulus properties could not be necessary to activate the neuron. Responses were

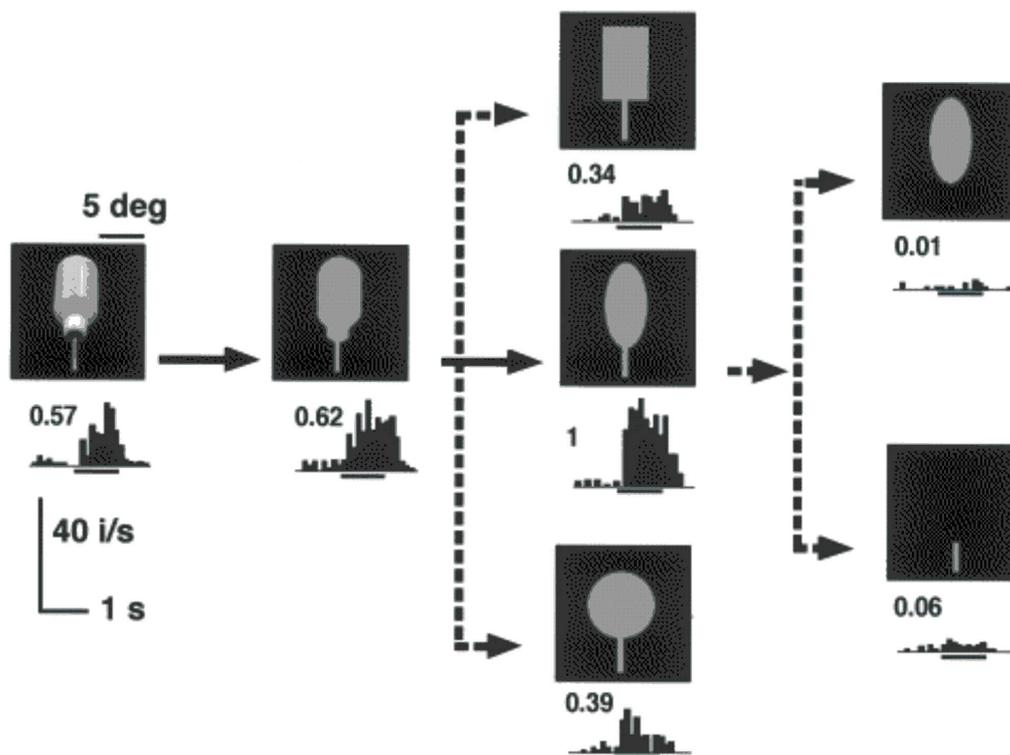


Figure 1.11: Exemplar simplification of a water bottle to determine the critical feature for a single TE neuron. Simplification proceeds from left to right. For each step, the stimulus image and the PSTH for a TE neuron are shown. PSTH are calculated from ten presentations of the stimuli. Numbers above the histograms represent the response magnitude, normalized to the magnitude of the largest response. In this case, the critical feature is an ellipse with a bar at the bottom (Wang et al., 1998).

considered to be equally strong if the mean firing rate did not differ significantly between the stimuli. The simplest 2-D feature that fully activated the cell was taken as the critical feature for the cell. An example of how simplification may proceed is shown in Figure 1.11.

Using this reduction technique, it was confirmed that most neurons in TE required complex stimuli to be maximally activated (Tanaka et al., 1991; Kobatake & Tanaka, 1994). In addition, the authors claimed to have identified a columnar organization in TE, with neurons in a column sharing the same critical feature (Fujita et al., 1992). Tanaka (1993) formalized their results as a feature-based shape representation in TE, in which neuronal columns signal the presence or absence of a specific feature in a shape. This theory can explain the broad selectivity of TE neurons, and resembles the columnar representation demonstrated for orientation in V1. The model is however limited by a missing rule how to derive the critical features, which are

specified only at a purely descriptive level. More importantly, there is some concern regarding the experimental methods. As described, simplification was continued as long as a neuron's firing rate remained maximal. The ability to detect deviations from the maximal firing rate – and thus changes introduced by a simplification – is however limited by ceiling effects. Nonlinear effects in general pose potential problems for the method. As an example, TE neurons have been shown to be excited by a certain object part if presented in isolation, but to be inhibited by the same object part if presented in combination with another stimulus (Tsunoda et al., 2001). Effects like this might have excluded objects from further testing even though they contained parts that would have activated the neuron. Finally, the way in which stimuli were decomposed into simpler parts, and thus the essential determinant for the critical features, seems questionable. The extraction of simpler parts was guided by the intuition of the experimenter only. Judging from the exemplary figures presented in the papers by Tanaka et al. (see examples in Wang et al., 1998; Tanaka, 2000), objects seem to have been decomposed into parts appearing to be “good” elementary shapes for a human observer. It is not obvious that TE decomposes objects in the same way.

A reductive approach as the one followed by Tanaka et al. is not the only possibility to study the selectivity of TE neurons. Constructive methods have been applied as well. These rely on the definition of a model for shape representation, from which parametric stimulus descriptions can be derived. Testing the responses of TE neurons to stimuli that systematically differ in these parameters allows to assess whether the selected shape parameters are encoded in TE. One such model of shape representation is the method of Fourier Descriptors, tested against the properties of TE neurons by Schwartz et al. (1983) and Albright & Gross (1990). The model describes shape boundaries only, parameterizing them by a set of so called Fourier Descriptors (FDs). FDs are computed by first determining the boundary orientation function for a shape – i.e. the orientation of the shape's boundary measured at regular intervals around the perimeter. This boundary orientation function is then expanded in a Fourier series, whose terms are the FDs. Each FD is thus associated with a specific frequency, amplitude, and phase, which correspond to the boundary's number of lobes, the lobe indentation, and the orientation of the closed contour. Any shape is fully described by a linear combination of FDs, with the additional advantage that this shape description is independent of position and size of the shape. Furthermore, neurons in early visual areas are known to be selective for the orientation of local contours, and therefore the boundary orientation function is an easily accessible information for TE neurons. In the model, each TE neuron would represent how strongly a particular FD is present in a shape.

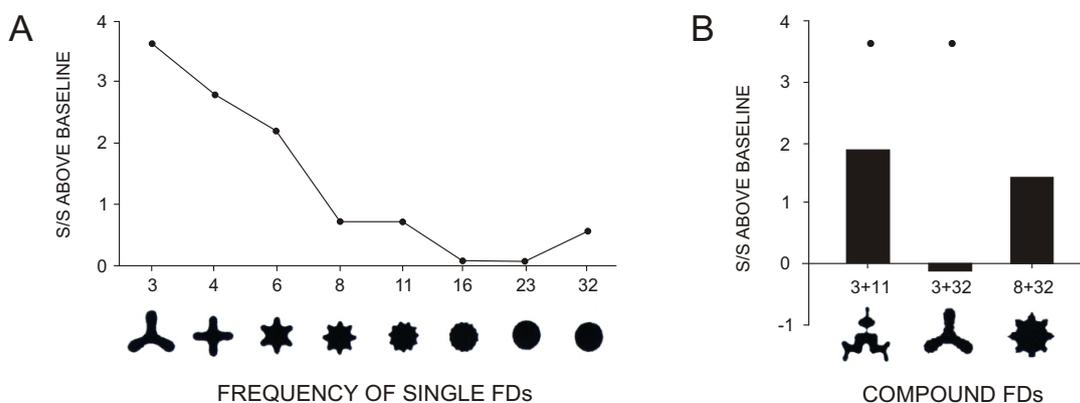


Figure 1.12: *A*, Selectivity of a single IT neuron for FD frequency. *B*, Actual (bars) and predicted responses (dots) of the same neuron for compound FDs. Responses to compound stimuli were predicted by summing the responses to single FDs, as assumed by the model. The neuron fails to extract its preferred FD when presented in conjunction with a second FD. Figures below the plots show the stimuli. S/S: Spikes per second (adapted from Albright & Gross (1990)).

Stimuli were constructed by inverse transforms of specific FDs. By using stimuli that were the inverse of a single FD, it could be demonstrated that about half of the recorded visually responsive TE neurons were systematically tuned to FD frequency. The response magnitude changed with the amplitude, but the frequency tuning curves remained similar. It therefore seemed possible that TE neurons encode shapes in terms of their FDs. However, contradictory results were obtained when stimuli were constructed as the inverse of the sum of two FDs. Under these conditions, neurons failed to respond to stimuli that contained their preferred FD, demonstrating that the two FDs were not represented independently as required by the model's assumptions (see Figure 1.12 for an example). In conclusion, the experiments could not demonstrate that the shape selectivity of TE neurons is generated by the contributions of specific FDs to the shapes.

The experiments described so far share the assumption that a shape is represented in the exact same way whenever it is encountered, independent of how it is perceived, and of which task an animal has to perform on the shape. Similarly, the learning history of the animal with a shape is not taken into account. No clear picture of the stimulus encoding properties of TE neurons has yet emerged from these experiments. However, this may change with experiments that explicitly test the effects of behavior or learning on the selectivity of TE neurons. It has long been suggested that behavior has an influence on TE neurons, possibly gating, potentiating, or even

altering their visual responses (Gross et al., 1979). Modulations of firing rate with task have been reported (Spitzer & Richmond, 1991). In addition, the way in which a shape was perceived, and not its physical contour, was recently demonstrated to determine the responses in TE (Baylis & Driver, 2001).

A number of recent experiments reported task or learning effects on the selectivity of TE neurons. In the first experiment, Sigala & Logothetis (2002) studied the subordinate classification of a set of face or fish stimuli. Monkeys were trained to categorize line drawings of faces or fish into two categories (exemplar face stimuli are shown in Figure 1.13). Shapes of both types were parameterized by four varying features (e.g. in the case of the face stimuli, eye height, eye separation, mouth height, and nose length). Category boundaries were defined based on the values of two diagnostic features (for the face stimuli, eye height and eye separation). Of the neurons that responded differentially to the values of at least one feature, a large percentage was selective for one or both of the diagnostic features, but not for the non-diagnostic features. The same effects could be obtained on the level of the whole population of visually responsive neurons. The population average for all neurons tested with the face, depicted in Figure 1.13, clearly shows the different selectivity for diagnostic versus non-diagnostic features. In conclusion, TE selectivity seemed to be shaped by the most task relevant subset of stimulus features.

Similar effects of training on the selectivity of TE neurons were found in a recent experiment by Baker et al. (2002). In this study, monkeys learned to discriminate among baton stimuli consisting of distinct top and bottom elements joined by a vertical stem. Stimuli were organized into tetrads representing the four possible combinations of two top and two bottom parts (see Figure 1.14 for stimulus examples). Monkeys were trained to respond to the members of two tetrads by pulling one of two levers, with the batons assigned to the response alternatives such that a decision could not be based on the identity of single parts. Instead, the monkeys had to take the conjunction of features into account to perform above chance. After completion of training, the activity of TE neurons was recorded during a fixation task in which the learned and similar, unlearned stimuli were presented to the monkeys. The authors observed a slight increase in the selectivity for learned versus unlearned stimuli. This increase in selectivity was analyzed more closely for each neuron by comparing the responses evoked by the members of a tetrad. By this means, a neuron's selectivity could be attributed to selectivity for batons' parts or for whole batons. In this data, the increase in selectivity for learned batons was evident in a modest enhancement of the selectivity for individual parts, and in a marked enhancement for the combination of parts. Since in the discrimination task the monkeys had to base their decisions on the conjunction of parts, these results

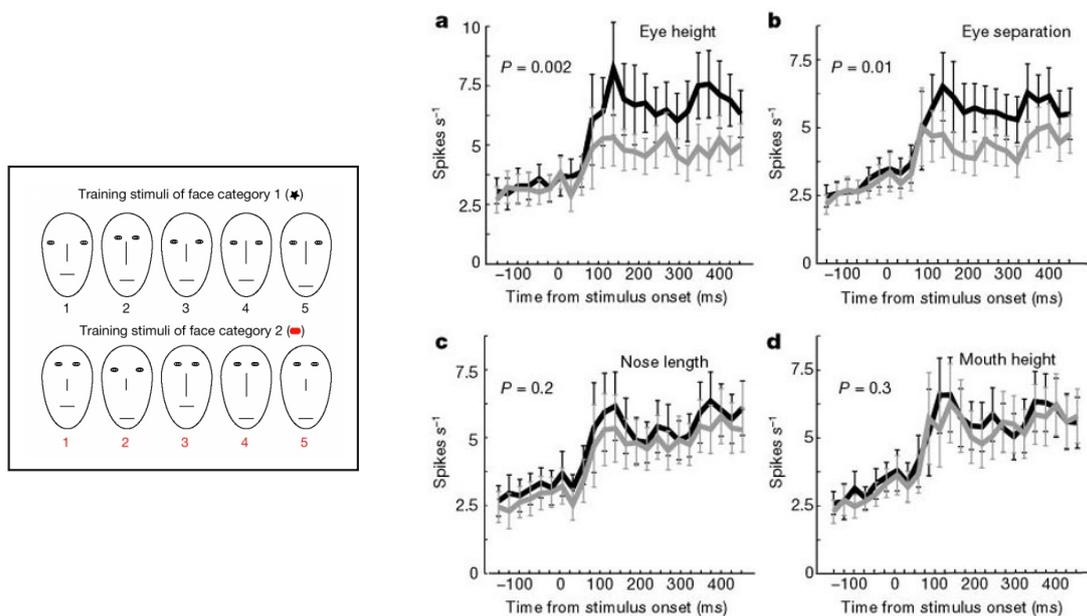


Figure 1.13: Stimuli and results of Sigala & Logothetis (2002). Left, Exemplar face stimuli, showing members of both categories. Right, Population average for the neurons tested with face stimuli, with data of each neuron normalized to its background activity preceding stimulus onset. Black traces indicate the average responses to the best feature value, and gray lines to the worst. Eye height and eye separation are diagnostic features.

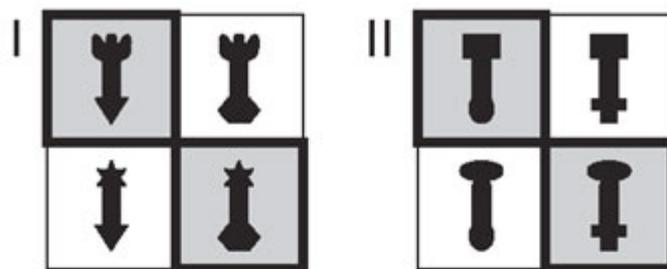


Figure 1.14: *Baton stimuli used by Baker et al. (2002). Each monkey learned to associate the members of two tetrads to two levers. Here, shapes with gray background were mapped to the left lever, and shapes with a white background to the right lever.*

again indicate that task requirements and learning interact with the selectivity observed in TE.

Another experiment from which conclusions about the impact of a task may be drawn is a study by Sheinberg & Logothetis (2001). In this experiment, monkeys were familiarized with 70 objects and were trained to respond to each of them by pulling one of two levers. Once the monkeys mastered this task, objects were either presented in isolation (isolated condition), or embedded in one of 100 randomly chosen natural scenes (embedded condition). When embedded in a scene, objects were located at a random position and were blended into the background scenes to increase the difficulty of locating the targets. In both conditions, the monkeys' task was to respond to the presented object by pulling the correct lever, which in the embedded condition required them to scan the background for any familiar object. No information about object identity was available before the trial started. Background scenes were large, complex images, and the monkeys were free to move their eyes throughout the trial. Recordings were carried out in TE while the monkeys performed trials of both conditions. Selectivity of an isolated unit was first established in the isolated condition. Units were then tested with effective and ineffective objects in the embedded condition. Responses were similar for isolated and embedded condition, suggesting that the observed activity was related to the process of noticing particular targets, independent of how they were found. In the embedded condition, neurons began to respond about 100 ms before the eyes acquired the target, but only if the monkey was about to fixate the target. Most importantly, very little discharge activity was observed while the monkeys explored the scene prior to locating the target, obvious in both examples shown in Figure 1.15. This was the case even though the neurons were visually responsive, and the backgrounds were rich scenes with a number of objects falling in a neuron's receptive field during the

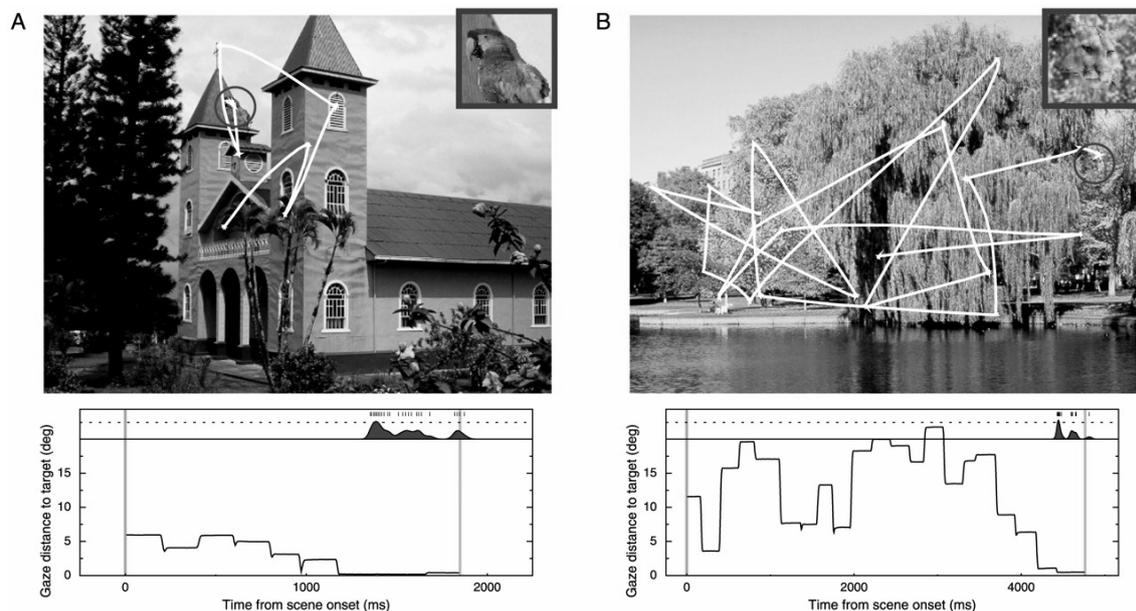


Figure 1.15: Behavior of TE neurons during a visual search task. A and B show two different stimuli of the embedded condition of Sheinberg & Logothetis as presented to the monkeys. Superimposed on the scenes are recorded eye movement traces. The red circle encompasses the position of the target, shown enlarged in the small inset. Plots below the images depict the gaze distance from the target (black trace) and the observed spike density function of a single neuron (red curve) as a function of time. The green line indicates stimulus onset, the blue line response time (Sheinberg & Logothetis, 2001).

visual search. Thus, although neurons appeared to have a broad selectivity which made them respond to a number of stimuli, they were sharply tuned to meet the requirements of the task by responding to the targets only.

In line with the conclusions of the above mentioned studies are the results of an experiment by Logothetis et al. (1995), which suggested that TE neurons may develop selectivity for complex stimulus configurations as the animals are trained to recognize specific objects. Monkeys were extensively trained to identify novel three-dimensional objects from two different classes. One class contained wire-like objects, the other amoeboid objects. The monkeys learned to identify these objects from particular viewpoints. After training, the responses of TE neurons to the trained objects, as well as to a large number of distractor objects, were recorded. A number of neurons were found which selectively responded to certain views of an object, but not to the distractors. Interestingly, when testing objects that the monkey could only identify from a certain viewpoint, no selective responses were ever encountered for those views that the animal failed to recognize. In addition, the number of cells

selective for a particular object class correlated with the amount of training that each animal had received for the object class. Taken together, the findings suggest that the neuronal selectivity was generated because of the training experience of the monkeys.

The findings of two further studies corroborate these conclusions. Sakai & Miyashita (1994) and Kobatake et al. (1998) repeatedly exposed monkeys to members of a fixed set of stimuli. In both cases, neurons were recorded in TE after extensive training. Manipulating the fine structure of the trained stimuli, Sakai & Miyashita (1994) observed that responses of TE neurons were always lower for the transformed than for the trained stimuli. Kobatake et al. (1998) compared the responses in the TE cortex of trained monkeys to the ones found in untrained monkeys. In the trained monkeys, more cells responded to the training stimuli than in the control monkeys. Also, TE neurons in the trained monkeys displayed higher selectivity for the training stimuli than the TE neurons in the control monkeys. Taken together, these studies strongly suggest that the selectivity of TE neurons is shaped by learning.

1.2.7 Effects of occlusion on the responses of TE neurons

In Section 1.1.2, the behavior of V1 and V2 neurons to partially occluded stimuli was reported. Here, the results of recordings in TE will be described. Kovács et al. (1995) compared the responses of TE neurons evoked by whole stimuli and by partially occluded versions thereof. Stimuli were simple geometric shapes, and the occluders were randomly placed masks, designed to occlude different amounts of these shapes (from 20 to 90%). For an example of these stimuli, see Figure 1.16A. Neural firing rate to occluded stimuli was reduced overall, an effect that increased with increasing amounts of occlusion. To further assess the influences of occlusion on the neural responses, stimuli were ranked for each neuron according to the mean firing rate that they evoked. Mean firing rate averaged across a population of neurons was then analyzed as a function of stimulus rank, giving a measure of the population selectivity (see Figure 1.16B). As long as sufficient stimulus information was visible through the masks, the dependency of firing rate on stimulus rank remained similar. This demonstrates that the selectivity of neurons remained unaffected. Most importantly, for all but the highest occluder density, mean firing rate was significantly higher for the best than for the worst stimulus, allowing these stimuli to be distinguished by their evoked firing rate. In conclusion, the results suggest that occlusion only weakly affects responses in the inferotemporal cortex.

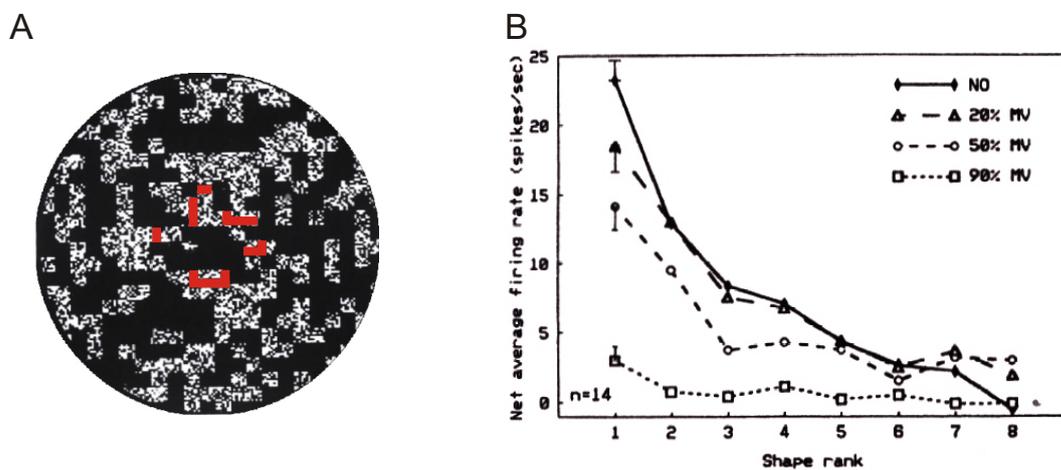


Figure 1.16: Influence of partial occlusion on the responses of neurons in the inferotemporal cortex. *A*, Exemplar stimulus as used in the study by Kovács et al.. The occluded shape is drawn in red for illustrative purposes only. *B*, Mean firing rate of a population of neurons as a function of stimulus rank. Shape 1 corresponds to a neuron's best shape, 8 to the worst shape, and the curve gives an indication of the neural selectivity. The full line shows the results for shapes presented without an occluder, dashed lines represent increasing amounts of occlusion (percentages give the occluder density). Firing rates are decreased with occlusion, but selectivity is mostly preserved (adapted from Kovács et al. (1995)).

An experiment by Missal et al. (1997) addressed the related question of how the responses of TE neurons changed if shapes were not presented in isolation, but overlapping a larger background. Although this study did not directly test the influence of occlusion since responses to the foreground object were investigated, it nonetheless allows to draw conclusions about possible interactions between an object and an occluder. The results show that shapes presented in isolation, and the same shapes presented in front of a larger background, could lead to markedly different responses. This was the case even though the shapes were perceptually segmentable for the monkey. The most commonly observed interaction between background and foreground shape was a reduction in firing rate. This parallels the reduction in firing rate for partially occluded shapes that has been described by Kovács et al. (1995). However, more complex, nonlinear interactions were also possible. Still, the neural responses were at least partially determined by the foreground shape: Some portion of the original selectivity for isolated shapes was preserved even if the shape was shown in front of a background shape. This remaining selectivity could completely be abolished if the shape and the background were painted in the same color, so that the foreground object could no longer be segregated from the background. The nonlinear effects therefore seemed not to be due to a failure to extract the foreground shape, but rather because of an interaction of the two shapes, indicating that in TE, responses to complex shapes cannot necessarily be predicted from the responses to the shape's parts. Although no similar effects were observed by Kovács et al. (1995), it is nonetheless possible that with different, maybe smaller masks, the interactions between masks and occluded shapes become more nonlinear.

1.3 Aim of the thesis

Since occlusion is ubiquitous in nature, it is important for an observer – either human or animal – to be able to identify an object even if only parts of it are visible. As described in Section 1.1, humans possess this capability fully at the age of one. In addition, most animals can identify objects correctly under partial occlusion conditions. However, it has been demonstrated for humans and for pigeons that there are limits to the recognition of partially occluded images. These studies have shown that it is a function of the spatial placement of an occluder whether an object can be identified despite partial occlusion. If the spatial placement of an occluder indeed has different effects on the behavior, then the question arises whether these effects are also reflected in the neural responses to a partially occluded stimulus. Specifically, it is interesting to test whether neural responses distinguish between partially occluded stimuli that can or cannot be identified. This is the main topic of this work. Not only does this question address how the brain deals with a situation that is very common in nature; since behavioral and neural performance are analyzed at the same time, it also addresses how much brain activity and behavior change in synchrony.

Chances are highest to detect spatial influences of occlusion if responses of individual neurons can be followed. This resolution is currently only possible by performing single cell recordings in the cortex of awake, behaving Rhesus monkeys. For Rhesus monkeys, it has not been tested so far whether the recognition of partially occluded objects depends on the spatial placement of an occluder. It has however been shown that monkeys performing a task find certain stimulus features more informative than others (e.g. Sigala et al., 2002; D’Amato & van Sant, 1988). Since information is therefore distributed inhomogeneously across a stimulus, occlusion will most likely have differential effects depending on the occluder placement. Rhesus monkeys thus are suitable subjects for the experiment.

From the experiments described in the introduction, it seems likely that occlusion has some effects on the whole visual cortex. However, the emphasis of the Ph.D. project is on how occlusion interacts with the encoding of complex objects. Because of its preference for complex shapes, area TE thus seems a good starting place for the experiments. Furthermore, it has already been demonstrated that TE neurons are affected to a certain degree by occlusion (Kovács et al., 1995). Yet, this study does not allow to determine whether the occlusion effects depend on where the occluder is placed.

In summary, the aim of this Ph.D. thesis is to test how different placements of an occluder affect both the behavior of Rhesus monkeys, as well as the encoding of partially occluded stimuli in area TE. With respect to the stimuli, the study takes advantage of the fact that occlusion can be used in conjunction with almost any stimulus. Most studies of TE responses have used line-drawings or geometric objects, because these simpler shapes can easily be parameterized. However, they do not represent the natural input to the visual system. While normally the complexity of a natural scene poses problems, this is not the case for the current study. To the contrary, the fact that natural scenes contain structure at many spatial scales is a benefit for the experiments, because it leaves a lot of room for interactions between occluders and images. The stimuli for the Ph.D. project will therefore be sets of natural scenes.

The complete experimental paradigm will be as follows: Monkeys are initially taught to discriminate sets of three natural scenes. The effect of partial occlusion on their behavior is then tested using the Bubbles paradigm developed by Gosselin & Schyns (2001). As described in Section 1.1.1, this paradigm involves showing stimuli behind occluders with randomly placed windows. The monkeys continue to identify the occluded scenes. The subsequent analysis compares the stimulus material from correct and incorrect trials. By these means, regions in the scenes can be identified where occlusion systematically generates identification failures (“diagnostic regions”). Similarly, scene regions where occlusion has no effect are identified (“non-diagnostic regions”). Human observers are tested under the exact same conditions, so that the distribution of information across natural scenes can be compared between monkeys and humans. The monkeys’ behavioral results are subsequently used to derive stimuli for single cell recordings. While recording from area TE, the natural scenes are either presented unoccluded or with occluders placed over the diagnostic or non-diagnostic regions. Occlusion of diagnostic regions renders the scene unidentifiable for the the monkeys, while occlusion of non-diagnostic regions does not affect the identification. By comparing the responses of TE neurons in these different conditions, the effect of occluder placement on the neuronal responses can be studied in a controlled way.

The proposed experiments become possible because it matters for the identification of a partially occluded stimulus where the occluder is placed. This is a consequence of the fact that different image regions carry different amounts of information. The experiments therefore not only test how occluder placement influences the perception of Rhesus monkeys or the encoding of occluded scenes. Additional conclusions can be drawn from the data. Based on the behavioral results, it is possible to identify which information Rhesus monkeys use to perform a task. While previous studies

on the categorization of natural scenes had to manipulate their stimuli to identify which features could be underlying a monkey's decision (see e.g. Vogels, 1999), we can determine without further manipulations which image regions are diagnostic. In addition, the same paradigm can be used for humans and monkeys, allowing a direct comparison of the behavior of the two species. On the neural level, the presentation of partially occluded images is equivalent to presenting only a selected portion of the image. The experiments by Tanaka have shown that neurons responding to complex objects are often also responsive to some parts of these objects (see Section 1.2.6). Here, we can test whether the same is the case for natural scenes, so that neurons responsive to the whole scene remain responsive for some portions of the scenes. Furthermore, since natural scenes in our study are split into diagnostic and non-diagnostic regions, we can address how the diagnosticity of a scene region influences its encoding, enhancing our understanding of how diagnosticity contributes to the responses in area TE.

Chapter 2

Methods

2.1 Animal preparation

G00 and B98, two adult male Rhesus monkeys (*Macaca mulatta*), weighing 10 and 14 kg, participated in the experiments. All studies were approved by the local authorities and were in full compliance with the guidelines of the European Community (EUVD, European Union directive 86/609/EEC) for the care and use of laboratory animals.

The monkeys were first trained to move calmly from their home cage into primate chairs, and they were familiarized with the setup environment. After this initial training, a custom designed titanium head post (Max Planck Institute for Biological Cybernetics, Tübingen, Germany), and a scleral search coil for eye position monitoring (Judge et al., 1980) were implanted in a sterile surgery. The following protocol is typically used in the lab for these surgeries: Animals are first premedicated with Glycopyrrolat (Robinul, 0.01 mg/kg), and Ketamine Hydrochloride (Ketavet, 15 mg/kg). Subsequently, anesthesia is induced with Fentanyl (Fentanyl-dihydrogencitrat, 3 μ g/kg) and Thiopental (Trapanal, 5 mg/kg). To maintain the anesthesia, Isoflurane is administered at levels of surgical tolerance (about 2%), while the animal is artificially ventilated. Throughout the surgical procedure, blood pressure, heart rate, and the respiration in terms of SpO₂ and CO₂ levels are constantly monitored. Body temperature is measured, and kept at 37 - 38 degrees using a heating pad. Postoperatively, the monkeys are daily administered an analgesic (Finadyne, 1 mg/kg) for three days, in addition to an antibiotic (Baytril, 5 mg/kg or Synulox, 0.1 ml/kg) for up to ten days.

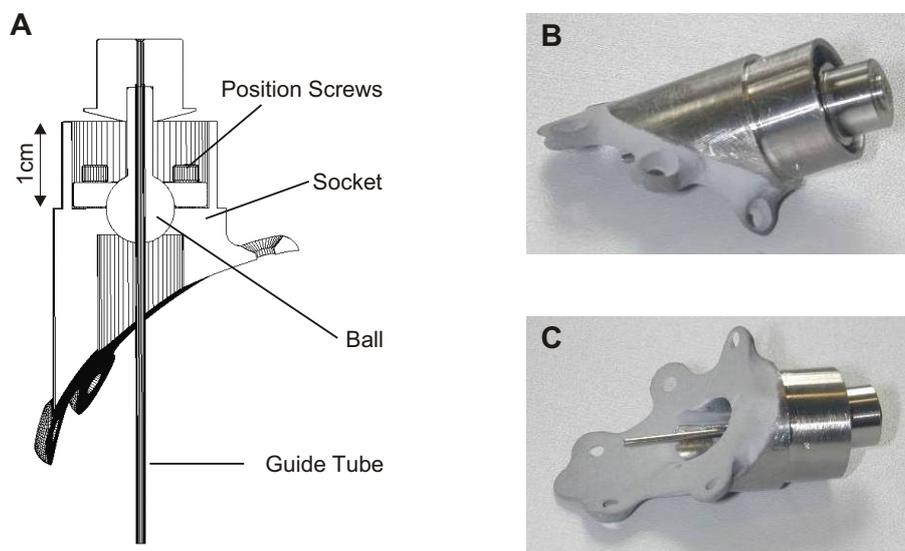


Figure 2.1: *Ball-and-socket chamber implanted on monkey G00. A, Construction of the chamber. The guide tube is held in place by the ball. Its angle can be adjusted by means of two position screws. The angle could be changed by pm5 deg from the central position. B and C, Views of the actual chamber. In C, the guide tube is visible.*

After sufficient behavioral data were collected from a monkey, a chronic titanium chamber for the electrophysiological recordings was implanted under sterile surgical conditions. The chamber (see Figure 2.1) consisted of a ball-and-socket joint with a stainless steel guide tube (1.3 mm outer diameter) passing through the chamber's center (Schiller & Koerner, 1971). The guide tube could be swiveled around the central position by loosening the position screws on top of the chamber. This allowed flexible positioning of the guide tube. To generate space for the guide tube to move, the chamber's base contained a large opening. It was filled with silicon (Sylgard 516 from Dow Corning GmbH, Wiesbaden, Germany; mixed in a ratio of A:B=1:1.2, and annealed at 150°C for 60 min). The silicon allowed the guide tube to move, but nonetheless tightly sealed the complete opening. With the exception of the recording sessions, the guide tube was closed with a stylus. The stylus was made from a metal tube slightly longer than the guide tube. One end of the stylus was bent so that it easily could be pulled out. The other end was closed by a drop of melted glass.

High-resolution magnetic resonance imaging (MRI) scans (3D-MDEFT, TR=22 ms, TE=4 ms, voxel size $0.5 \times 0.5 \times 0.5 \text{ mm}^3$, 160 slices) of each monkey's head and brain were obtained at the beginning of the experiments. These served to design

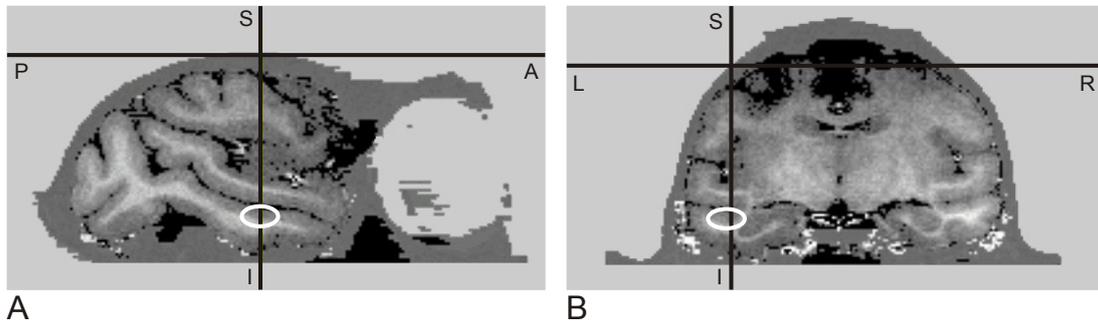


Figure 2.2: *Two cross-sections of the brain of monkey G00, obtained from the structural-anatomical MRI data. A, Sagittal view. B, Coronal view. The yellow ellipse encloses the intended recording location in the anterior lower bank of the STS. Red lines intersect at the chamber position on the skull. Abbreviations: I - inferior, S - superior, A - anterior, P - posterior, L - left, R - right.*

the chamber with optimal fit to each monkey's skull. MRI scans were furthermore used to place the chamber on the skull such that the intended recording region could optimally be reached. To determine the appropriate chamber position for a monkey, suitable for recordings in TE, the anterior half of the STS was first located in the brain scans. The yellow ellipse in Figure 2.2 shows the recording region on the brain anatomy of monkey G00. Starting from the recording position, the electrode track was determined next. During this procedure, care had to be taken that electrodes avoided critical brain regions. For the TE recordings, this was the middle cerebral artery in the bank of the lateral fissure. Extrapolation from the recording position along the electrode track to the skull determined the chamber position. Once its position had been defined, the chamber base was exactly modeled to the surface of the skull at this position. The precise fit between chamber and skull minimized the risk of infections. Horsley-Clark coordinates for the chamber position were AP 15.4, ML 16.8 for monkey G00, and AP 18.1, ML 17.7 for monkey B98 (AP - anterior-posterior, ML - medial-lateral).

2.2 Monkey psychophysics

2.2.1 Stimuli

In each experiment, monkeys were exposed to a stimulus set of three images. During large parts of the experiments, the stimuli were shown behind occluders generated using the Bubbles paradigm. Stimulus sets will be described in the next section, before explaining the construction of the masks.

Stimulus sets

Three stimulus sets were tested, one of them showing simple geometrical shapes (G1 - G3), and two containing natural scenes (N1 - N3, N5 - N7). All stimulus sets are shown in Figure 2.3. The geometrical shapes were generated in CorelDraw (www.corel.com). They had a size of 256 by 256 pixels, and were constructed by placing a small square on one of three sides of a larger square. Both squares were painted light on a dark background.

All natural scenes were taken from Corel PhotoCDs. Images again had a size of 256 by 256 pixels. They were normalized to have equal Fourier amplitude spectra. This was achieved by computing the Fourier transform of each image, resulting in an amplitude and phase spectrum for each image. Amplitude spectra were then averaged over the set members. Finally, the average amplitude spectrum was combined with the image specific phase spectrum in the inverse Fourier transform to yield a normalized image. The overall image contrast was reduced by limiting image gray-scale values to the range from 30 to 70% of the maximal gray scale value. If shown without further scaling, images of a size of 256 by 256 pixels covered 6 by 6 deg of visual angle¹.

Generation of masks

Masks were constructed as suggested by Gosselin & Schyns (2001). A masked image appeared to be occluded by a surface punctured by round windows (“bubbles”), through which parts of the image were visible. This effect was generated by blending between an image and a mask as large as the image. At each image pixel, the mask specified a value between 0 and 1, expressing the percentage of original image trans-

¹In the remainder of the text, deg will be used to refer to degrees of visual angle.

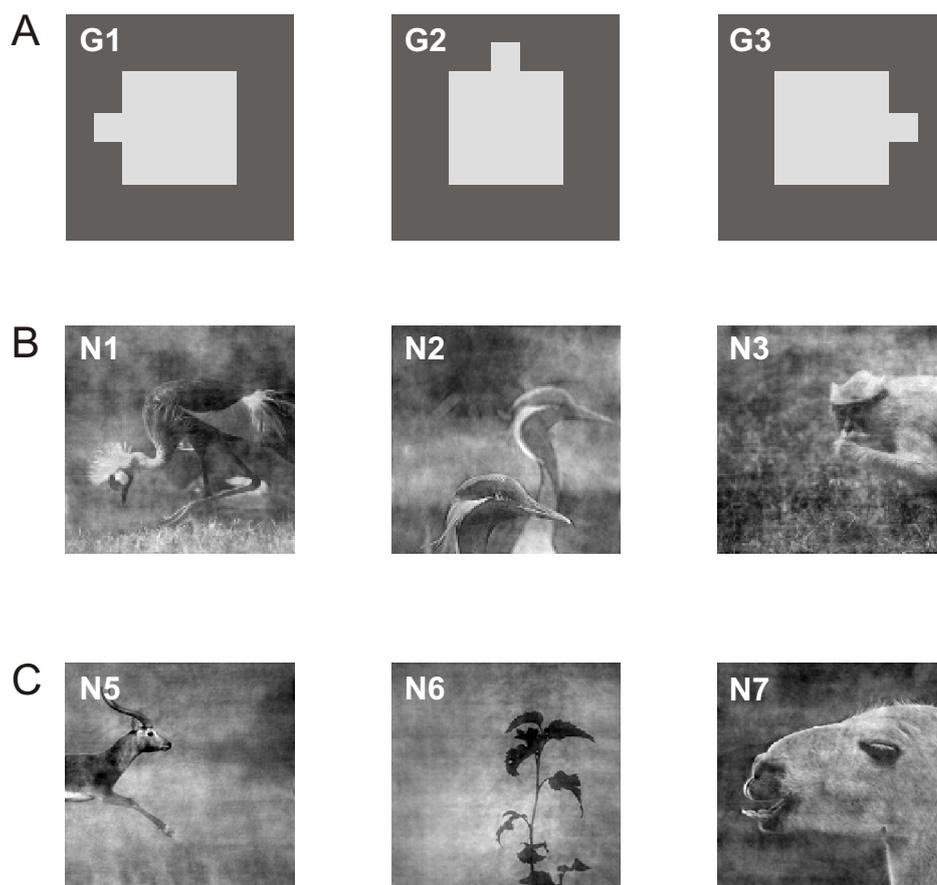


Figure 2.3: *Stimulus sets of the Bubbles experiments with monkey observers. A, Geometrical patterns. B and C, Natural scenes. Small letters in the stimuli are the stimulus labels used for referencing. They were not part of the actual stimuli.*

mitted through the mask. Computationally, blending was performed by calculating the intensity value

$$o_i = m_i \cdot u_i + (1 - m_i) \cdot b$$

at each pixel i , where o denotes the gray-scale value of the occluded image, u the gray-scale value of the unoccluded original image, m the mask value, and b the background gray-scale value.

To generate masks with round windows, all mask values were initially set to zero, generating a nontransparent surface of the background color. 2-D Gaussian profiles with a peak value of 1 were then added to this surface. Each Gaussian profile corresponded to one bubble in the mask. Profiles were computed as

$$f(x, y) = \exp\left(-\frac{(x - x_c)^2 + (y - y_c)^2}{\sigma}\right).$$

x and y give the horizontal and vertical coordinates of each image pixel, respectively. x_c and y_c are the coordinates of the maximum for the 2-D profile. σ sets the profile's width. Because of their Gaussian envelope, windows smoothly merged into the nontransparent background. Bubbles could be manipulated by the parameters defining the 2-D Gaussian curve, i.e. center coordinates and width. The center coordinates determined the position of a bubble, while the width set its size. Each image pixel could be selected as a center position. No bubble could therefore be centered around a location outside of the image. However, since center coordinates could fall onto or close to the edge of the image, bubbles only partially overlapping the image were possible.

Masks with multiple bubbles were generated by adding multiple 2-D Gaussian profiles. The composite mask was computed as

$$M(x, y) = \sum_{i=1}^N \exp\left(-\frac{(x - x_{ci})^2 + (y - y_{ci})^2}{\sigma}\right),$$

where x and y again specify the horizontal and vertical coordinates of the pixels. N denotes the number of bubbles, and x_{ci} and y_{ci} are the center coordinates of the i th bubble. Gaussian profiles were restricted to have different center positions at least in one direction. Bubbles could therefore overlap, but could never fall completely on top of each other. Because of the blending operation between mask and image, mask values were restricted to the range from 0 to 1. Addition of several overlapping Gaussians could in principle lead to mask values greater than 1. Therefore, after addition of all Gaussians, mask values were clipped to 1. An

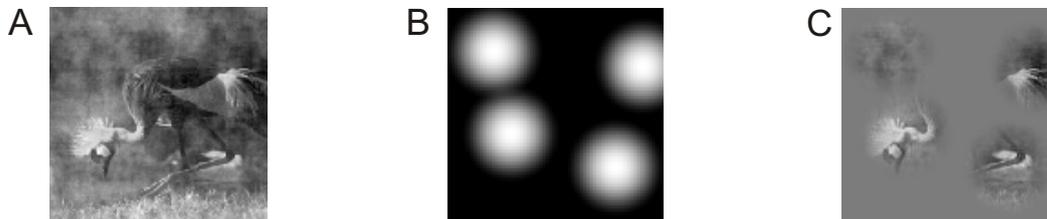


Figure 2.4: *Construction of masked stimuli. A, Original image. B, Exemplar mask containing four bubbles. Brighter values indicate higher transparency. In the actual experiments, smaller bubbles were used. C, Image in A covered with the mask in B.*

example for the generation of masked images is given in Figure 2.4, showing the original image, a mask, and the masked image.

2.2.2 Experimental paradigms

Most of the behavioral experiments used a stimulus discrimination task, which will be described in detail in the next section. Some changes were necessary to incorporate the Bubbles paradigm. These will be described thereafter. To compare the distribution of information across a scene as determined with Bubbles to the one found during visual examination of the same scene, two variants of a free viewing task were devised. The first tested the eye movements elicited by a mere inspection of a scene without the context of a task, while the second addressed the influences of a task on this data. Both tasks are described below.

Discrimination task

A trial began with the presentation of a yellow fixation spot in the center of the screen, combined with the sounding of a tone. The monkey had 4 s to acquire fixation on this spot. 100 ms after the onset of fixation, the spot was turned off and one of the stimuli appeared centrally for 300 ms, after which the fixation spot was again presented for 100 ms. Stimuli were shown with their original size of 6 by 6 deg. When the fixation spot was finally turned off, three small white squares (the targets) were presented in the periphery. All targets had a distance of 6 deg from the center of the screen, and the monkey was given 4 s to make a saccade to one of them. Foveation of the selected target for 300 ms ended the trial. A high-pitched

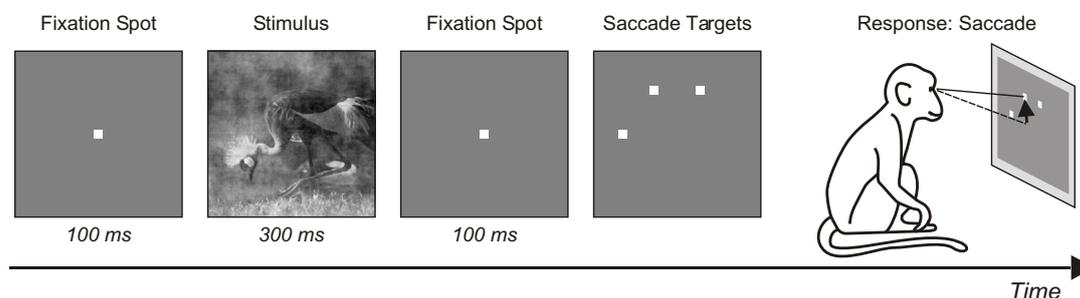


Figure 2.5: A typical trial during the discrimination task performed by the monkeys. Trials began with the onset of a fixation spot, followed by the stimulus. They ended with the presentation of three peripheral response targets, one of which the monkey had to select by fixating on it. See text for more details.

tone was played, and the screen went blank. The next trial was initiated after an inter-trial interval (ITI) of 3 s. Figure 2.5 illustrates a typical trial for this task.

From the acquisition of fixation on the centrally presented fixation spot until the onset of the targets in the periphery, the monkey had to keep his gaze within a window of 2 deg radius around the center of the screen. He was then allowed to move his gaze freely from the central position to one of the target positions. However, as soon as one of the targets was foveated, the monkey's gaze had to remain in a 2 deg window around this position. If at any point during which fixation was required the animal's gaze left the fixation window, the trial was aborted. The same happened if the monkey failed to acquire fixation within the set times at the beginning of the trial, or during the response period. If a trial was aborted, the screen turned blank immediately, and the ITI was extended by an additional delay period of 2.5 s. The long delay was intended to discourage the monkey from aborting trials.

Stimulus sets consisted of three images. Each image was associated with one of the response targets, so that each of the three stimuli required a unique response from the monkey. Only fixation on the correct target was rewarded by a drop of juice at the end of the trial. Incorrect responses were not rewarded, and were additionally indicated by the sounding of a specific tone at the end of the trial. A pseudo-random sequence was used to determine which stimulus was presented in a trial to assure that each stimulus was used equally often. On every trial, the number of presentations was determined for each stimulus. The stimulus with the lowest number of presentations was shown on the next trial. If multiple stimuli had been shown equally often, one of them was randomly selected.

When a new stimulus set was introduced, the monkeys had to be taught the associations between stimuli and corresponding response target locations. This was done by introducing a brightness difference between the response targets. The correct target was always shown as a white square, while the brightness of the incorrect targets was reduced. During the first trials with a new stimulus set, the brightness of the incorrect targets was adjusted so that they disappeared against the screen background. Brightness was then slowly increased so that the performance of the monkey always remained above 70% correct. Training was completed when the monkeys could perform the task better than 80% correct with all targets at the same brightness. This training procedure took several weeks for the first stimulus set. For subsequently introduced stimulus sets, the training duration could be reduced to about a week.

Bubbles task

The monkeys were first trained to perform the basic discrimination task on a stimulus set before the Bubbles task was introduced. Order and timing of stimulus presentations were kept identical for the two tasks, as well as the response mode: As before, the monkeys responded to the stimulus by making a saccade to one of the targets. The only difference between the two tasks therefore was the presentation of occluded stimuli, which were only used in the Bubbles task. Presentations of masked and unmasked stimuli were interleaved. At least every third trial contained a stimulus without a mask, to be able to control the monkey's behavior in the basic task. The same pseudo-random procedure as described above was used to determine the stimulus identities for a trial. Monkeys were rewarded in the case of both masked and unmasked stimulus presentations when making a saccade to the correct target.

As described in Section 2.2.1, masks were non-transparent surfaces punctured by randomly placed bubbles. Each mask could be characterized by the number of bubbles, their size, and positions. Size was equal for all bubbles, and kept constant throughout the experiments. Bubbles were generated from a 2-D Gaussian with a width σ of 20 pixels. This resulted in moderately large windows (1.35 deg in diameter for stimuli that were 6 by 6 deg). Bubble position was randomly determined on every trial. The number of bubbles in a mask determined how much of a stimulus was visible, and therefore set the overall difficulty of the task. Since the monkeys had to remain motivated, the task could not be made too difficult, which required a flexible setting of the bubbles numbers. Three different methods were used interchangeably:

- *Constant bubble numbers:* The number of windows was kept constant for a large number of trials, eventually reduced by the experimenter. The primary use of this scheme was during the first introduction of the Bubbles task, with the goal of familiarizing the monkeys with the presentation of masked stimuli. In this situation, stimuli had to be easily identifiable despite the masks, and large, constant numbers of bubbles were required.
- *Staircase protocol:* The number of bubbles was adjusted by a staircase protocol. Staircases were updated to maintain a performance level of 75% correct. Every fourth presentation of a stimulus, the performance of the monkey with this particular stimulus was determined for the last four trials. Unmasked presentations were not taken into account. If all four last responses had been correct, the number of windows was reduced by three, while for three correct responses the number of windows was kept constant. If two or less responses were correct, the bubble number was increased by two. Bubble numbers were adjusted individually for each stimulus.
- *Method of constant stimuli:* In this case, upper and lower limits were set for the bubble numbers. Bubble numbers started with the upper limit, and were reduced by a fixed amount after a small number of trials (usually 4 or 5 trials). When the bubble number reached the lower limit, it was reset to the upper limit, to be reduced again. The upper limit was usually chosen so that the task was very simple for the monkey, while performance at the lower limit was at threshold. With this method, masks for different stimuli had the same number of windows.

In a few additional experiments, fixed masks were used instead of the randomly varying Bubbles masks. These experiments were designed to verify the Bubbles results, and they were carried out after the Bubbles testing sessions were concluded. When these masks were used, two changes had to be introduced in the Bubbles paradigm: The ratio between masked and unmasked trials was changed, so that more trials contained unmasked than masked trials. In these experiments, only every third trial contained a masked stimulus. In addition, responses on masked trials were randomly rewarded in 50% of the trials, irrespective of the correctness of an answer.

Eye movements tasks

A set of three tasks was used to study the eye movements elicited by different stimuli. The first paradigm was designed to give a measure of the precision with which monkeys could fixate on any location on the screen. In this calibration task,

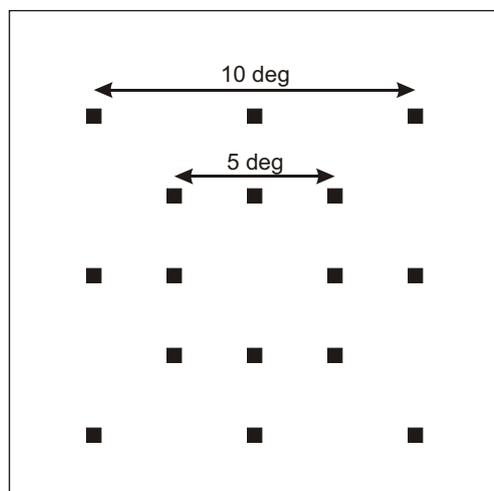


Figure 2.6: *Screen positions for which the fixation accuracy was tested. See text for task description.*

the monkeys had to fixate on a white square that was presented at various positions on the screen. In every trial, the square initially appeared in the center of the screen. If the monkeys acquired and successfully held fixation on this position, the square jumped to one of 16 peripheral locations. The monkeys were required to reacquire fixation on this location to receive a drop of juice as reward. For every position, a fixation was counted as successful if the monkey held its gaze within a circle of 2 deg diameter around the position for 400 ms. If these requirements were not met, a trial was aborted and no reward was given. The 16 positions were located on the corners and the midpoints of the sides of two concentric squares (see Figure 2.6). The larger square had a side length of 10 deg, the smaller one of 5 deg. Thus, fixation accuracy was determined for horizontal and vertical distances up to 5 deg from the center of the screen.

A free viewing task was used to study where monkeys looked on an image when it was presented to them for visual exploration. The intention of this task was to study the natural viewing behavior of a monkey without the influences of a task. Therefore, no constraint was placed on where the monkeys had to look, and no reward was given to them. Images had a size of 10 by 10 deg, and were presented for 3 s. During an experimental session, the images from one of the stimulus sets were used. They were shown to the monkey in a pseudo-random order. Stimulus presentation cycled through the stimulus set as long as the monkey was willing to look at the images.

In the third eye movement task, it was tested whether the looking pattern evoked by a scene changed when the monkeys had to give a response to the image instead of merely inspecting it. So called “scanning trials” were therefore introduced in the standard discrimination task. These scanning trials had a longer stimulus presentation time of 3 s than normal trials. No fixation requirements were imposed while the stimulus was visible. The rest of the trial remained unchanged with respect to the normal discrimination paradigm. Central fixation was required before and after stimulus presentation. Also, as usually the monkeys had to give a response to the scenes by making a saccade to one of the response targets. In both trial types, stimulus size was 6 by 6 deg. Scanning trials were indicated to the monkey by showing a large frame around the first fixation spot. During a session, every third or fourth trial was a scanning trial, while the rest were normal trials as in the original task.

2.2.3 Apparatus

Behavioral training was performed in a specially designed awake monkey physiology setup. During any procedure in this setup, monkeys were seated in a custom-made primate chair (Max Planck Institute for Biological Cybernetics, Tübingen, Germany). These chairs were placed within an electromagnetically and acoustically shielded chamber. During experimental sessions, the light in the chamber was turned off. Monkeys could be monitored by the experimenter by means of two infrared cameras. While in the setup, the monkey’s head was fixated.

Eye movements were monitored using the scleral search coil technique (Robinson, 1963). In the initial surgery, a search coil was implanted in one of the monkey’s eyes. In the setup, monkeys were placed in a high frequency magnetic field, inducing a voltage in the search coil. Since the amplitude of the induced voltage is proportional to the sine of the angle between the axis of the search coil and the magnetic field, it is a measure of the animal’s gaze direction. Two sets of field coils were used to generate a horizontal and a vertical magnetic field in the chamber. The two fields alternated at different frequencies. Separately detecting voltage components at the two frequencies allowed to determine horizontal and vertical components of the eye position. With the available equipment (CNC Engineering, Seattle, Washington), eye positions could be measured stably and with a high precision. Data points were taken at a rate of 200 Hz.

Stimuli were presented on a 21” monitor (Intergraph 21sd115, Intergraph Systems, Huntsville, USA) with a resolution of 1024 by 768 pixels, and a refresh rate of 75 Hz. Background luminance of the monitor was set to 41 cd/m², and the monitor was

gamma corrected. The monitor was placed outside the electromagnetic shielding at a distance of 95 cm from the monkey. For monitoring, stimuli were also simultaneously presented on a second computer screen in front of the experimenter.

Stimuli were generated and displayed with STIM (D.L. Sheinberg), executed on a dedicated PC. STIM provides a TCL/TK interface into OpenGL. A STIM script was used to load images, to generate random masks, and to combine the two into masked images by blending. In addition, STIM was also responsible for saving the masks. Each mask was saved as a compressed binary file to reduce the required amount of storage. While STIM was used for stimulus presentation, the execution of an experiment was controlled by a so called experimental state system in the QPCS environment. The state system was programmed in C, and implemented on a distributed network of PCs, all of which used the real-time operating system QNX (QSSL, Ontario, Canada). The state system timed stimulus presentations, and additionally provided the behavioral control for the experiments. It registered the eye position and, if applicable, tested it against predefined fixation criteria. By these means, trials in which the monkey failed to fixate were detected and aborted. By analyzing the eye position, the state system could also determine which response target the monkey selected in the discrimination and Bubbles task. Furthermore, the state system determined which stimulus to show on a given trial, and when appropriate set the mask parameters. All events in a trial were written to a file. Dedicated FastEthernet served for the communication between all setup computers.

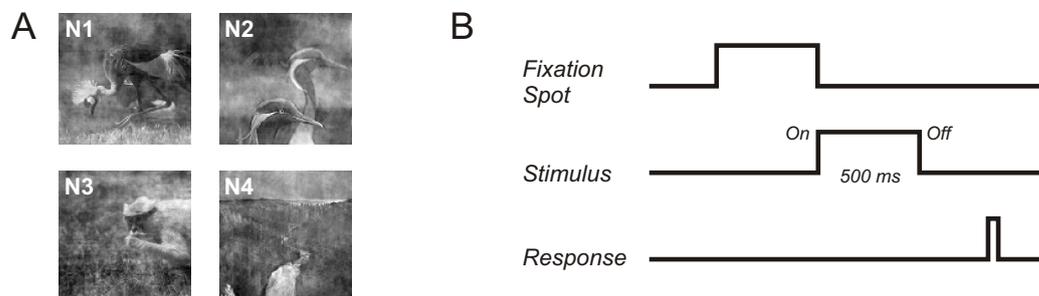


Figure 2.7: *A, Extended natural scene set 1. B, Time sequence of the task used with human observers. The upper two lines indicate the time course of the appearance of fixation spot and stimulus on the screen. The lowest line shows the time course of the response (see text for a more thorough description of the paradigm).*

2.3 Human psychophysics

2.3.1 Subjects

A total of 12 subjects participated in the described experiments. In most cases, three subjects were tested per experiment. All subjects were naïve as to the purpose of the experiments. Subjects had normal or corrected-to-normal vision. Testing sessions usually lasted two to three hours, with subjects completing between 1100 and 2000 trials in this time. Subjects returned to the lab for additional sessions, until a total of 4000 to 6000 trials had been collected.

2.3.2 Stimuli

Human observers were only tested with the two sets of natural scenes. The second set (scenes N5 - N7) was identical for human and monkey observers, while the first set included scene N4 as an additional fourth scene (see Figure 2.7). Images from the first set were also presented at different sizes; in the initial experiment, they were shown with a size of 128 by 128 pixels. In later experiments, they had a size of 256 by 256 pixels. During the experiments, the distance of the subjects from the monitor was such that the smaller images subtended 4.2 deg, and the larger images 8.3 deg.

2.3.3 Experimental paradigm

Trials began with the presentation of a yellow fixation spot for 500 ms, followed by one of the experimental stimuli for 500 ms (see Figure 2.7 for illustration of the task). Observers responded after the presentation of the stimulus by pressing designated keys on the numerical keypad of a standard computer keyboard. No constraints were imposed on reaction time. After a subject's response, the next trial was initiated.

Each of the images in a stimulus set was associated with a specific response key. Subjects were first familiarized with the stimulus set, and learned to press the correct keys in response to the images. Once they had correctly performed 20 trials with non-occluded stimuli, the actual testing began, in which stimuli were shown behind masks partially revealing the images. As before, observers had to report the identity of the presented stimulus. In trials in which they were unsure about the image identity, they were instructed to nonetheless choose one of the response alternatives. No feedback was given about the correctness of their answer.

Masks were constructed as described in Section 2.2.1. As for the monkey observers, bubble size was kept constant throughout an experiment. With human observers, Gaussian profiles with a width of either 10 or 20 pixels were used, corresponding to about 0.95 or 1.9 deg, respectively. Bubble position was again randomly determined on every trial. Using a staircase protocol, the bubble number was set such that subjects identified a stimulus correctly in 75% of the trials. Staircases were run for each stimulus independently. Every fourth presentation of a stimulus, the bubble number was updated depending on the correctness of the last four presentations. It was decreased by three if the subject had responded correctly in all last four trials, and increased by two if less than three answers had been correct. For three correct responses, the bubble number remained unchanged. The stimulus presented in a trial was chosen pseudo-randomly as described in Section 2.2.2, assuring that every stimulus was shown equally often. Every tenth presentations of an image, it was shown without a mask. For each image, the presentation of unmasked images was started a random number of trials after the beginning of a testing session. Otherwise, the unmasked presentations of all images would have immediately followed each other in consecutive trials. Responses to unmasked images were not taken into account when updating the number of bubbles.

In some cases, subjects were presented with a questionnaire after they had completely finished an experiment. The questionnaire was intended to determine which decision strategies an observer was aware of. Questionnaires showed printouts of the experimental stimuli. The subjects were instructed to circle in these printouts the image regions they had found most informative in identifying the images. They

were also provided with additional space to add descriptions of any other strategy they used during the task.

2.3.4 Apparatus

Stimuli were presented on a 21" monitor (Intergraph 21sd107, Intergraph Systems, Huntsville, USA) with a refresh rate of 85 Hz and a resolution of 1152 by 864 pixels. The monitor's RGB channels were gamma corrected individually. Background luminance was set to 27 cd/m². Subjects were comfortably seated in front of the screen with an approximate viewing distance of 60 cm. However, the subject's head position was not controlled. A table in front of the subjects held a computer keyboard, which the subjects used for their responses.

Stimulus generation and presentation was controlled using STIM, executed on a dedicated graphics workstation. Besides stimulus generation, STIM was also responsible for saving the masks. Stimulus presentation was timed by a QPCS state system, implemented on a second PC operating under QNX. The state system additionally logged a number of events during each trial, including the subject's response. It also determined the content of the trials by designating the stimulus to be presented, in combination with computing the appropriate numbers of bubbles for a mask. Communication between QNX and STIM computers was by dedicated FastEthernet.

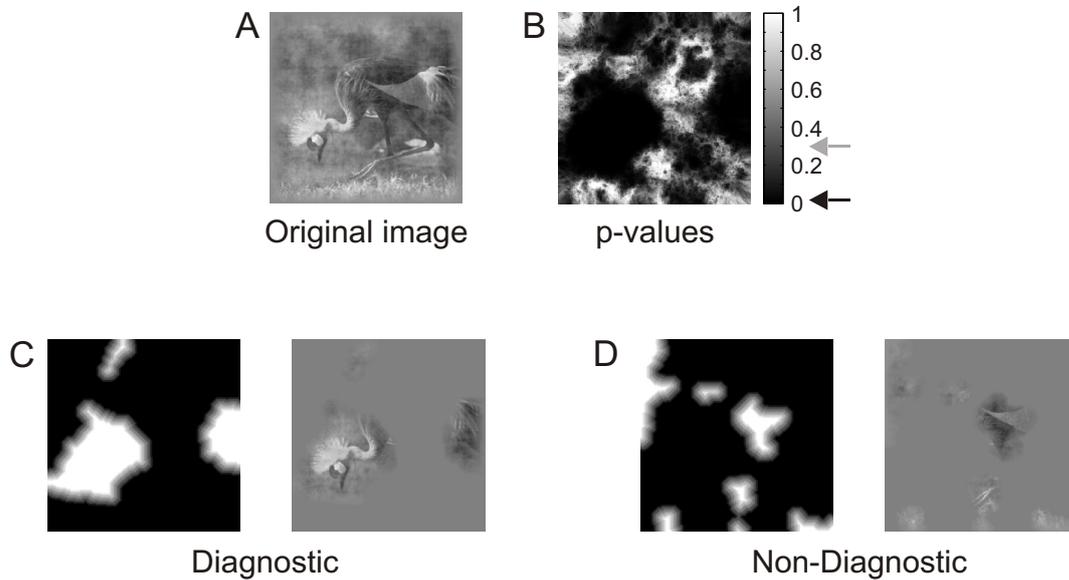


Figure 2.8: Construction of masks for the physiology stimuli. *A*, Original image. *B*, Map of p -values computed for monkey G00 as Bubbles result for this stimulus. The black arrow indicates the 30th percentile, while the gray arrow is placed at the 70th percentile. By selecting pixels with p -values below the 30th percentile, and applying the smoothing described in the text, the diagnostic mask shown in *C* is generated. Similarly, selecting pixels with p -values above the 70th percentile and smoothing leads to the non-diagnostic mask shown in *D*. *C* and *D* also show the result of overlaying the computed masks on the original image.

2.4 Neurophysiology

2.4.1 Stimuli

Depending on the task, different stimulus sets were used. In the fixation paradigm, we selected four of the six natural scenes used for the monkeys before. In the case of monkey G00, these were images N1, N3, N5, and N7. For monkey B98, we selected images N2, N3, N5, and N6. Images were shown without masks, and with masks derived from the Bubbles results. Masks were computed individually for each monkey.

As a result of the Bubbles task, a map of p -values was computed that specified for each image pixel whether masking of this pixel had a significant effect on behavior (see Section 2.5.1). Based on these maps, six different masks were constructed for each of the images (see Figure 2.8 for an illustration). Masks again contained

values between 0 and 1, and controlled the blending operation between image and background. Initially, all mask values were set to zero. Next, some proportion of the mask values was set to 1, making the underlying image region visible. Which of the mask values were set to 1 was based on the p-value map for an image. By selecting mask pixels where the corresponding p-values were at the lower end of the p-value distribution, masks were generated through which diagnostic image regions were visible (diagnostic conditions). On the other hand, masks with holes over non-diagnostic image parts were built by selecting mask pixels with p-values in the upper end of the distribution (non-diagnostic conditions). Three diagnostic and three non-diagnostic masks were constructed. In each group, the three masks differed by the amount of the underlying image that was visible (the visible stimulus size). This was achieved by setting either 10, 30, or 50% of the mask values to 1, so that either 10, 30, or 50% of the original image was visible. Mask computation was done using Matlab (www.mathworks.com).

Some of the generated masks contained very small elements. They also sometimes had very rough edges. This fine detail in the masks was unnecessary, and sometimes distracting. To smooth mask edges and remove small, disconnected elements, image dilation followed by image erosion operations were repeatedly applied to the masks. These operations were taken from the Matlab Image Processing Toolbox. Dilation added pixels to the mask boundaries by setting each pixel value to the maximum of the values in a defined neighborhood. In contrast, erosion removed mask pixels by setting pixel values to the minimum value in the neighborhood.

The resulting masks still contained only values of 0 or 1, and therefore had sharp transitions from mask to image. These edges were smoothed by replacing the sharp transitions from 0 to 1 at the edges with ramps in which mask values increased in steps of 0.1 from 0 to 1. As a side effect, this manipulation now blended the masked images smoothly into the screen background. However, presentation of the unmasked images generated sharp edges between image and background. To avoid differences between masked and unmasked images, a similar manipulation had therefore to be done for the unmasked images. Edges were here removed by generating a vignette for each image. The vignette was a 20 pixels wide frame with the same outer side lengths as the images. In the vignette, mask values were ramped from 0 at the outside to 1 in the inside in steps of 0.05.

Masked and unmasked images were also equalized in terms of their lower order image properties. Mean luminance was matched by setting the mean of all image pixel values to 0.5 in each image. Image contrast was adjusted for the masked images by scaling the image values so that their standard deviation was equal to the standard deviation of the image values in the unmasked image. In case of the masked images,

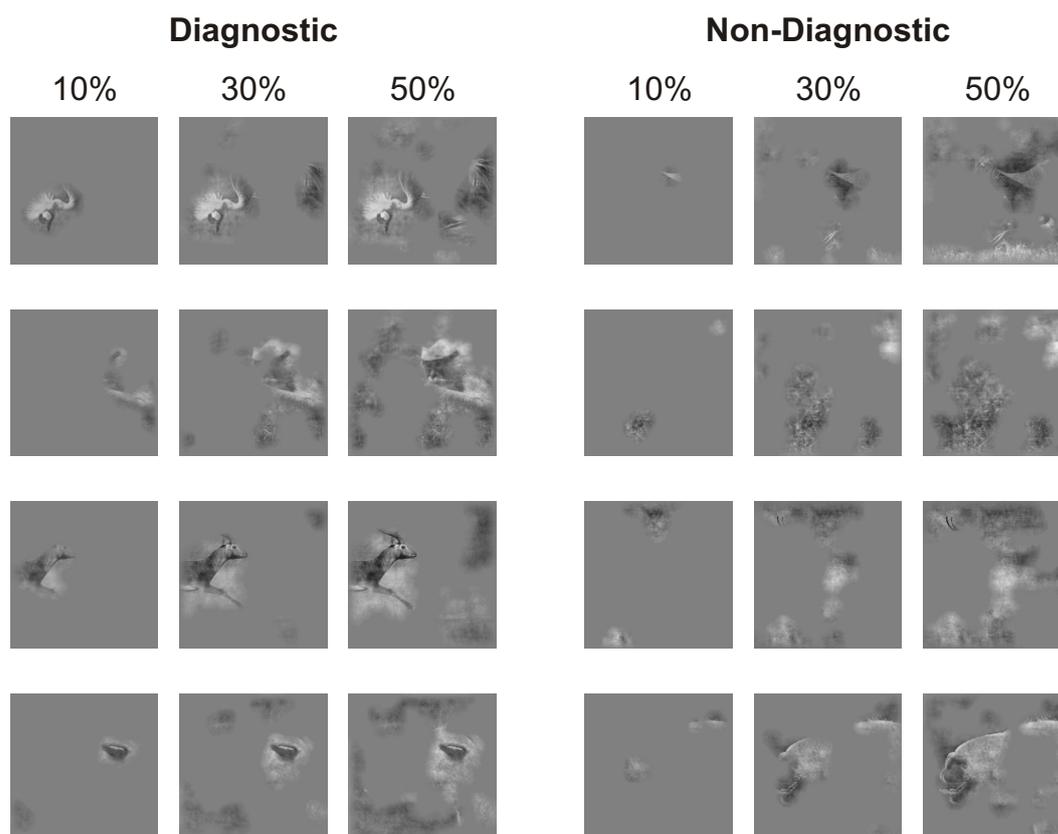


Figure 2.9: Stimulus set constructed for monkey G00. In the diagnostic conditions, diagnostic image regions were visible, while they were occluded in the non-diagnostic conditions. Percentages specify how much of the underlying image was visible through the mask.

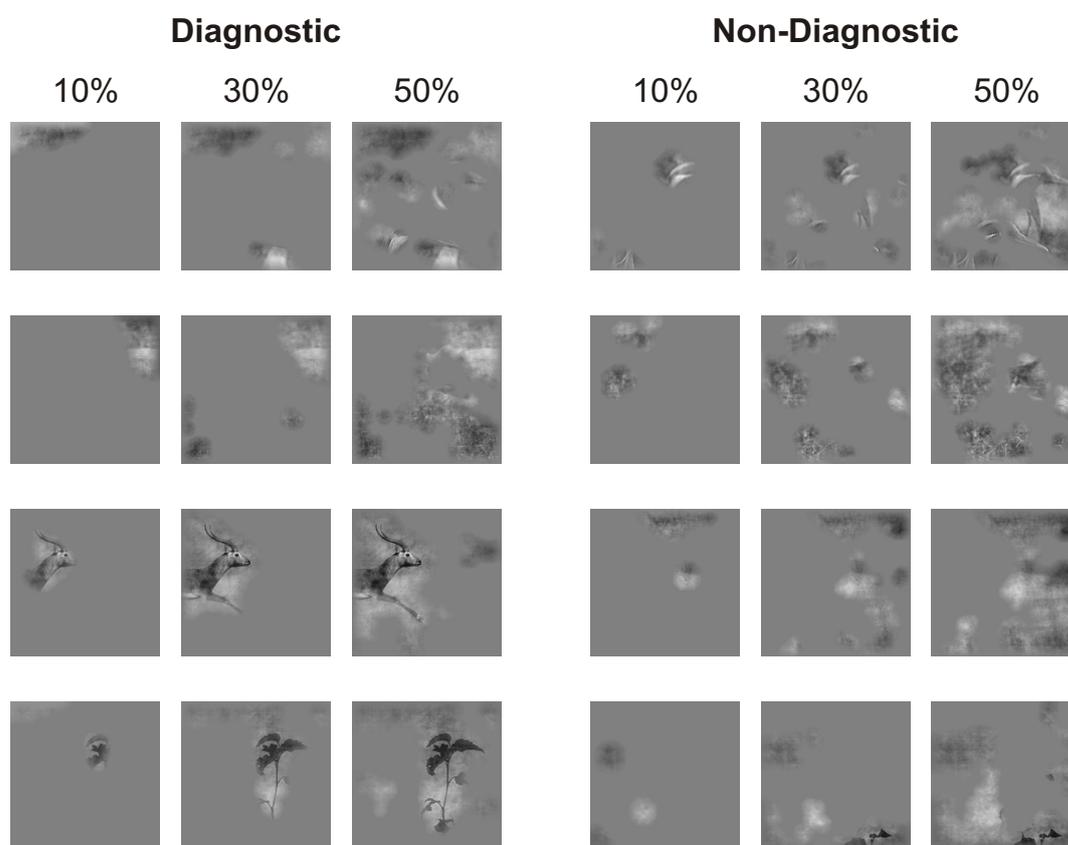


Figure 2.10: *Stimulus set constructed for monkey B98. See Figure 2.9 for more details.*

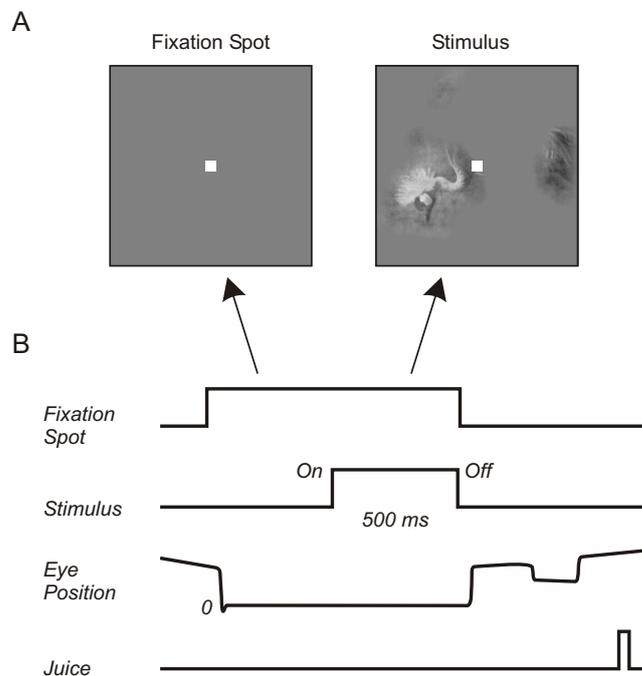


Figure 2.11: *Design of the task used in the physiology sessions. The sequence of events is outlined in B, while A shows exemplar stimuli appearing in the baseline and stimulus period. The arrows indicate during which time intervals these stimuli were shown. The upper two lines in B show the time course for the presentation of fixation spot and stimulus. The third line gives an exemplar eye movement trace. This line indicates the distance of the gaze from the center of the screen. Fixation at the center of the screen is reached at the level indicated by the small “0”. The lowest line, finally, shows at which time point juice was delivered to the monkey.*

only the values of visible image regions were taken into account when normalizing mean and standard deviation; masked regions were not considered. All stimuli were shown at a size of 6 by 6 deg. The stimulus sets for both monkeys are shown in Figures 2.9 and 2.10.

2.4.2 Task

Trials began with the presentation of a red square in the center of the screen (see Figure 2.11 for an illustration of the task sequence). Monkeys were additionally alerted to the beginning of a trial by the sounding of a tone. They were required to fixate on the square within 3 s after its appearance. If they acquired fixation

within this time limit, the stimulus appeared behind the red square some time after the onset of fixation. The stimulus was presented in the center of the screen. The duration between fixation onset and stimulus onset was at least 400 ms; it was however extended up to 600 ms depending on the performance of the monkey. Due to a problem in the experimental control system, the stimulus erroneously appeared 100 ms after fixation onset in a few sessions for monkey G00 and B98. This short presentation duration of the fixation spot may introduce an unstable baseline. Neurons may still be responding to the fixation spot when the stimulus is turned on. However, the fixation spot was present in all conditions. Response differences between conditions can thus not be an artefact of the too short baseline period, and the data of the sessions were included in the analysis.

Stimulus presentation lasted 500 ms, after which time the stimulus was removed. Throughout the whole period, the monkeys had to keep their gaze within a window of 1 deg diameter around the center of the screen. They were in principle rewarded for fixating the required amount of time, but to increase their attention, a second task was introduced. In most trials, the screen went blank after stimulus offset. Reward was delivered, if appropriate, 1 s thereafter. On some trials, however, the fixation spot jumped from its central position to one of four peripheral locations. The monkeys had to make a saccade to the new location to receive a reward. The next trial was initiated after an ITI of 3 s.

Trials were aborted for both monkeys if the fixation requirements were not met. For monkey B98, an additional sensor made it possible to record jaw movements. To reduce movement artifacts during the recordings, jaw movements were prohibited during the trial until the delivery of juice. The threshold for the jaw movements varied on every recording session, depending on the quality of the physiology recordings, as well as the position of the sensor with respect to the jaw. Thresholds were set such that jaw movements that generated detectable artifacts in the single unit recordings led to an abortion of the trial. As described before, stimulus sets for this experiment consisted of four natural scenes, presented in seven conditions. A pseudo-random sequence determined the combination of stimulus and condition presented on a trial. Between 10 and 20 successful repetitions were collected for each stimulus and condition.

2.4.3 Setup and recording procedure

Neural signals were recorded using a 5-channel electrode drive (5-channel Mini-Manipulator from Thomas Recording GmbH, www.ThomasRecording.de), which allowed

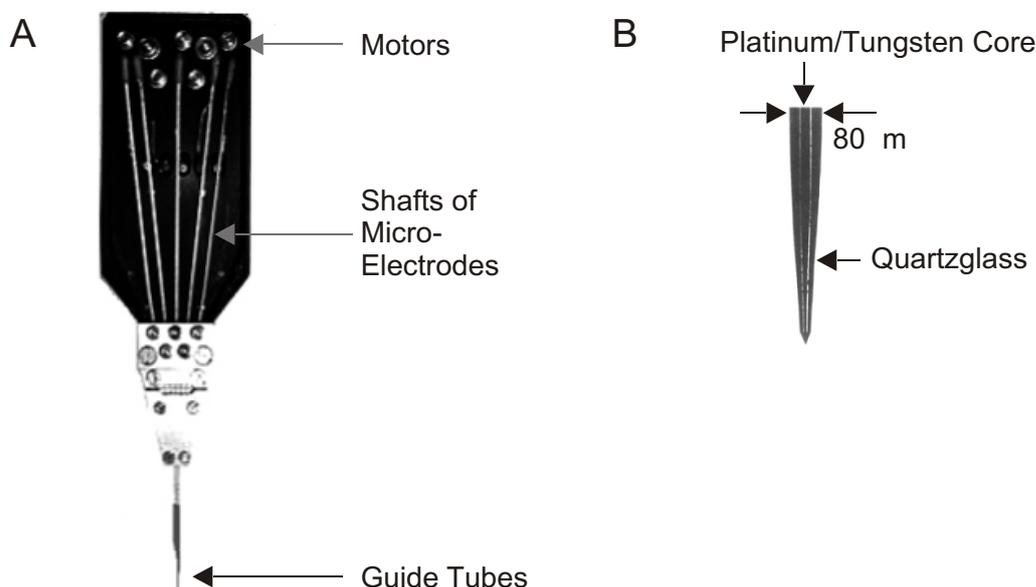


Figure 2.12: *A, 5-channel electrode drive from Thomas Recordings. Each electrode is fed through a metal guide tube. Guide tubes are bundled at the end of the drive. Electrodes can be moved independently by their own motor. B, Photo of an electrode tip, showing the Pt/T core and its Quartzglass cover.*

five electrodes to be advanced independently in $1\ \mu\text{m}$ steps. The drive and a sample electrode tip are shown in Figure 2.12. Electrodes consisted of a Pt/T core coated with Quartzglass, and had an impedance between 1 and $2\ \text{M}\Omega$ (ESI2ec, Thomas Recording GmbH). Each electrode was fed through a separate guide tube with an outer diameter of $305\ \mu\text{m}$ and a length of either 88 or 93 mm. These guide tubes were inserted into the brain through the chamber's guide tube.

Signals were amplified to span the range of $-5\ \text{V}$ to $5\ \text{V}$, and low-pass filtered with a cutoff of 3 kHz (MeasUnit5 Amplifier, Thomas Recording GmbH). The amplified signal was subsequently divided into multi-unit activity (MUA) and local field potential (LFP) by applying different filters. To generate the multi-unit signal, the amplified signal was band passed between 500 Hz and 10 kHz, while 1 Hz and 100 Hz were used as upper and lower cutoff frequencies for the LFP. Before saving the signals to disk, the recorded MUA and LFP were digitized with 22.3 kHz and 4.46 kHz, respectively.

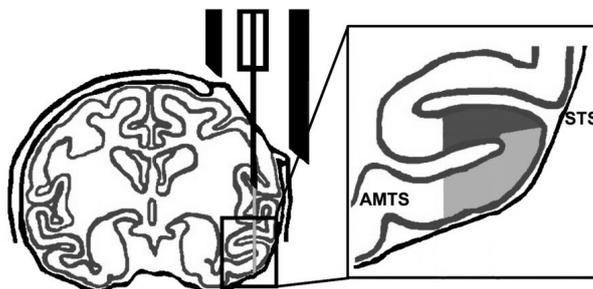


Figure 2.13: *Left, Coronal view of the brain with a schematic chamber, guide tube, and electrode. Right, Enlarged view of the STS region. The dark gray region corresponds to the lower bank of the STS, the light gray region to lateral TE (adapted from Janssen et al. (2000)).*

Electrode tips were lowered until they reached the lower bank of the the STS or lateral TE (see Fig 2.13), determining the recording depth by transitions between gray and white matter, as well as sulcal boundaries. Electrodes were then slowly advanced until at least one single unit could be isolated at each channel. For each channel, the first single unit that could be stably recorded was usually selected for further testing, irrespective of its responsiveness or selectivity to the experimental stimuli. This method was chosen to guarantee an unbiased sampling of TE neurons. For every electrode, the lower end of the STS and the recording location were noted.

The same setup chamber was used as before (see Section 2.2.3). For monkey B98, jaw movements were recorded in addition to the eye movements. Jaw movements were detected by an infrared motion sensor (AMN22112, www.nais-e.com) that was attached to the primate chair. The setup chamber provided electromagnetic shielding during the recording sessions. To further reduce artifacts in the recorded signals, the grounding scheme was adapted such that the setup chamber, the monkey, all amplifiers, and the shields of all cables leading into or out of the chamber were grounded to the same point. Computer control of the experiments was done with the same network of distributed PCs as described in Section 2.2.3. In addition to the network nodes used before, two PCs were used to record MUA and LFP. Start and stop of the recordings were triggered by the state system at the beginning and end of each trial.

2.5 Data analysis

2.5.1 Analysis of Bubbles data

After every masked trial of the Bubbles experiments, the used mask was saved, together with the subject's response, and the identity of the presented stimulus. As soon as all testing sessions were completed, trials were sorted according to which stimulus was presented, so that the rest of the analysis could be performed separately for each stimulus. In the next step, the correctness of the observer's response was determined for each trial of one stimulus. Based on the response, the masks were split into two groups, one containing the masks from all correct trials, the other one from all incorrect trials. The following analysis was then concerned with identifying differences between these two groups. The analysis of differences between masks proceeded pixelwise, i.e. by only taking mask values at one of the image pixels into account. The results of any of the analyses could then be summarized by combining the results obtained for all pixels, and representing the combination as an image.

Qualitative differences between the masks in correct and incorrect trials were analyzed by computing the average of each of the two groups. These average masks again contained values between 0 and 1, indicating how visible image regions were in an average correct or incorrect trial. In a quantitative approach, the mask value distributions in correct and incorrect trials were computed at each pixel. Distributions were then compared between correct and incorrect trials using the Kolmogorov-Smirnov test. Since the test was performed at each image pixel, the resulting p-values were Bonferoni corrected for the number of pixels.

To analyze data from single sessions, the correlation coefficient between the mask values at a pixel, and the correctness of the responses were computed following a suggestion by Chauvin et al. (submitted). To compute the correlation coefficient, correct trials were assigned a response value of 1, and incorrect trials a response value of 0. The correlation coefficient was then calculated as

$$C_i = \frac{\sum_t m_i(t) \cdot r(t)}{\sqrt{\sum_t (m_i(t) - \bar{m}_i)^2 \cdot \sum_t (r(t) - \bar{r})^2}},$$

where C is the correlation coefficient, $m(t)$ the mask value in trial t , $r(t)$ the response value in the same trial, \bar{m} the average mask value computed over all trials, and \bar{r} the average response. The subscript i indicates computation of a value at the i th

pixel. The correlation coefficients were additionally z-transformed to have a normal distribution:

$$Z_i = \frac{C_i - \bar{C}}{s_C}.$$

Here, \bar{C} indicates the average correlation coefficient over all pixels. Likewise, s_C is the standard deviation of the correlation coefficient over all pixels. All of the analysis was done using Matlab.

2.5.2 Construction of a model observer

A model observer was constructed to identify which diagnostic regions are generated if a particular response strategy is followed during the task. The model observer was designed to discriminate between the three geometric shapes G1 - G3. It used the following decision strategy to identify shapes: Shapes G1 and G2 were identified by the position of the small square in these shapes. Shape G3 was not directly encoded. It was identified whenever a shape could not be matched with either G1 or G2.

The model was implemented in the following way: Two templates were generated to detect G1 and G2. Each template was zero at all image pixels except for the pixels falling onto the small square of one of the shapes. In this region of 30 by 30 pixels, the template was set to 1 (see Figure 2.14 for an illustration of the model). These templates were applied to shapes G1 - G3. To simplify the computation, the gray-scale values of the stimuli were scaled such that all shape regions had a value of 1, and all background regions had a value of 0. When a stimulus was presented to the model observer, it computed the value of a decision variable d by multiplying the stimulus with the template at each pixel, and summing across the whole image. Values of d greater than 0 were only reached if template and stimulus matched. When tested repeatedly, observers usually identify a shape with a certain error rate even under optimal conditions. A noise term was therefore introduced in the computation of the decision variable. The noise was drawn from a Gaussian distribution with a mean of 0. By these means, the noise could increase or reduce the decision variable. In conclusion, the decision variable was computed as

$$d = \sum_i T_i \cdot S_i + r.$$

Here, the sum is carried out over all image pixels i . T_i and S_i indicate the template and stimulus value at a pixel, and r denotes the random noise term. d was computed for both templates. d_1 is used to refer to the value of d for the template matching G1, and d_2 for the template matching G2. Larger values of a decision variable indicate a

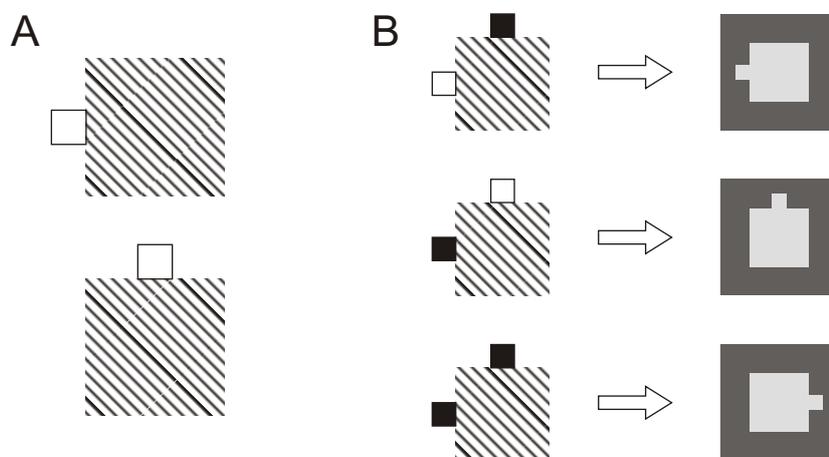


Figure 2.14: *Model observer. A, Model templates. The model has two detection windows, which are located at the positions of the small rectangles of G1 and G2. Outlined regions show the detection windows; the hatched regions correspond to each shape’s large rectangle, and are only shown to indicate the position of the detection windows. B, Model decision rules. Pixel values in the left window larger than a threshold, combined with pixel values smaller than a threshold in the upper window, generate “G1” responses. Similarly, G2 is detected by larger pixel values in the upper window, and values smaller than threshold in the left window. If the threshold is exceeded in none of the windows, the model responds “G3” (see text for a more thorough description).*

better match between template and stimulus. The model thus identified a presented stimulus with G1 if d_1 was above a certain threshold, and d_2 below the threshold. Similarly, the model chose G2 if d_2 was above and d_1 below threshold. Finally, the response G3 was selected if both d_1 and d_2 were below threshold. Both threshold and noise term could be adjusted so that the model’s performance matched that of a real observer. In the simulation run in this project, a threshold of 225 was chosen. The noise term was adjusted by setting the standard deviation of the Gaussian distribution to 240.

The model could be exposed to masked shapes as well. In this case, shapes were occluded with the Bubbles masks. The same decision variables were then computed for the masked shapes. Again, to simplify the computation, shapes were transformed so that their gray-scale values were either 0 or 1. Using the Bubbles masks, these shapes were blended into a black background, i.e. a background with a gray-scale value of 0 (see Section 2.2.1 for an explanation of the blending operation). This generated masked stimuli that also had gray-scale values between 0 and 1. The same decision rules were used as before to identify the occluded shapes. Because of

the occlusion, the model observer sometimes failed to identify a shape. The masks for which this was the case could be subjected to the same analysis as the data collected for human or monkey observers.

2.5.3 Eye movements

As a prerequisite for the analysis of the scanning data, the data collected during the calibration task was used to assess the precision of the fixation data. During each trial in the calibration task, the monkeys had to hold their gaze at one of several peripheral locations for 300 ms. Each location was tested several times. Thus, the variability in trials of the same location gave a measure of how much fixation positions were scattered around a location that the monkeys intended to fixate. To quantify the variability, the horizontal and vertical deviation between the required and the actual gaze position was computed on every trial. The actual gaze position was determined as the mean eye position within the 300 ms fixation interval. The absolute values of the differences were then averaged over all repetitions for a location, specifying the fixation accuracy as a function of the spatial position of a fixation location. By averaging over all spatial locations, a general mean error was computed for the horizontal and vertical direction.

The rest of the analysis was concerned with the scanning behavior evoked by natural scenes. Only fixation periods were analyzed, i.e. periods in which the eye position remained stable. These were identified as time periods of at least 100 ms with eye movement velocities below 60 deg/s. At a sampling rate of 200 Hz, this criterion corresponds to a spatial displacement in eye position of less than 0.3 deg between consecutive data points. The eye position within each fixation period was determined as the mean position during the fixation duration.

When fixation probabilities were computed for different image regions, the fixation positions identified in the free-viewing task were combined with the precision data obtained from the calibration task. Instead of assigning a fixation to a single position only, fixations were attributed to a rectangular region delineated by the determined fixation location plus and minus the fixation error in horizontal and vertical direction. By combining all fixations in this way, the number of fixations falling onto a specific image point could be computed across the whole image. Probabilities were generated by normalizing this data to the total number of fixations falling onto an image. The first fixation after any stimulus onset was excluded from the computation.

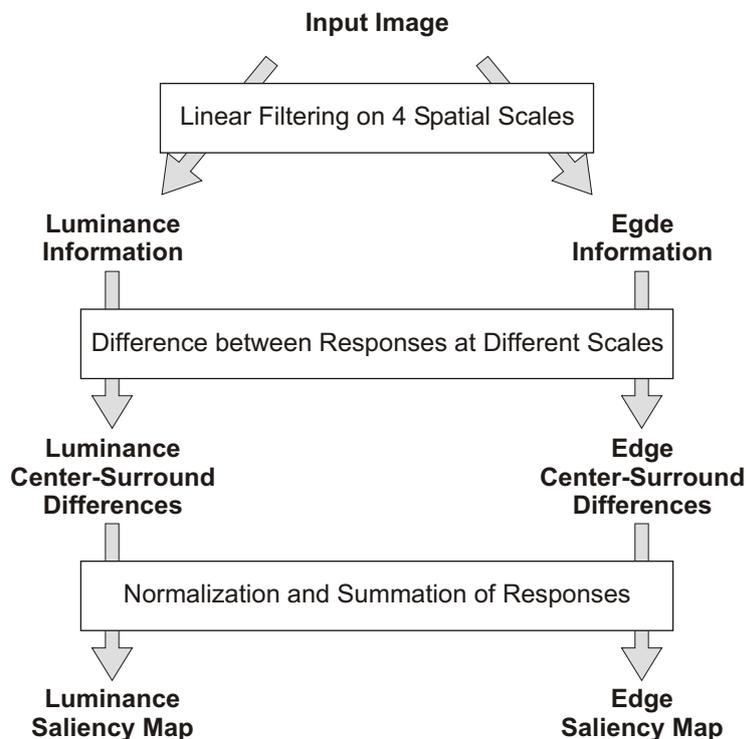


Figure 2.15: *Characterization of low-level properties of natural scenes. To compute luminance and edge saliency maps, several analysis steps were necessary, as described in the text. Further analyses use the results computed at the first and the last stage.*

2.5.4 Image characterization

To characterize the physical properties of the employed natural scenes, the distribution of luminance across the images, as well as the distribution of edge energy, was analyzed using Matlab. The full analysis progressed through multiple stages as shown in Figure 2.15, and followed a model proposed by Itti et al. (1998). The results from the first stage (luminance and edge information), as well as the luminance and saliency maps computed at the last stage were used for further description of the natural scenes.

Analysis at the first stage proceeded through four spatial scales, corresponding to four spatial resolutions. These were generated through progressively low-pass filtering and subsampling each image. At the finest resolution, images had their original size of 256 by 256 pixels. Each subsequent scale reduced the image size by a factor of 4, so that at scale 4, images had a size of 32 by 32 pixels. Luminance information was then computed at each of the four scales by convolution of the image with a

2-D Gaussian with a kernel size of 20 by 20 pixels and a standard deviation of 4 pixels. Edge energy was computed at four different orientations (0° , 45° , 90° , 135°) at each of the four spatial scales. It was extracted by applying oriented Gabor filter pairs to the images. Gabor filter pairs consisted of a symmetric and an asymmetric filter, which were constructed as the product of a sine or cosine function and a 2-D Gaussian envelope:

$$G_s = \cos(2\pi f x) \cdot \exp\left(\frac{x^2 + y^2}{2\sigma^2}\right),$$

$$G_a = \sin(2\pi f x) \cdot \exp\left(\frac{x^2 + y^2}{2\sigma^2}\right).$$

Here, x and y give the horizontal and vertical coordinates of an image pixel, σ the standard deviation, and f the spatial frequency of the filter. The standard deviation was set to 4 pixels, and the frequency to 1/10 pixels. Artifacts at the image borders were avoided by appending copies of an image to its borders. These copies were only present while convolutions were computed; they were removed after these calculations. All computed luminance and edge maps were rescaled to half the size of the original image, i.e. 128 by 128 pixels. In summary, at this stage of the analysis, four feature maps were computed indicating regions with high or low luminance in the images. 16 feature maps showed the distribution of edges at different orientations and different scales. Examples are given in Figure 2.16A and B.

In the next analysis steps, the derived multiscale feature maps were combined into so called saliency maps. To compute saliency maps, center-surround responses were first computed from the multiscale luminance and edge maps. A center-surround organization is akin to the visual receptive fields found for example in the retina, the geniculate nucleus, and the primary visual cortex. These neurons are most sensitive in a small region of the visual space (the center), while stimuli presented in a broader region concentric with the center (the surround) inhibit the neuronal response. In the model, center-surround interactions were implemented as differences between a fine and a coarse scale map for a feature. Scales 1 to 3 were selected as fine scale maps; the coarse scale maps had an offset of 1 to 3 scales from the fine scale. This meant that for example differences between maps at scale 1 and scale 2, 3, and 4 were computed. Two types of center-surround interactions were realized for the luminance information, representing On-Center or Off-Center responses. For On-Center responses, maps at the coarse scale were subtracted from the fine scale map, while for Off-Center responses, the fine scale map was subtracted from the coarse scale one. In both cases, the resulting differences were rectified. By this means, 12

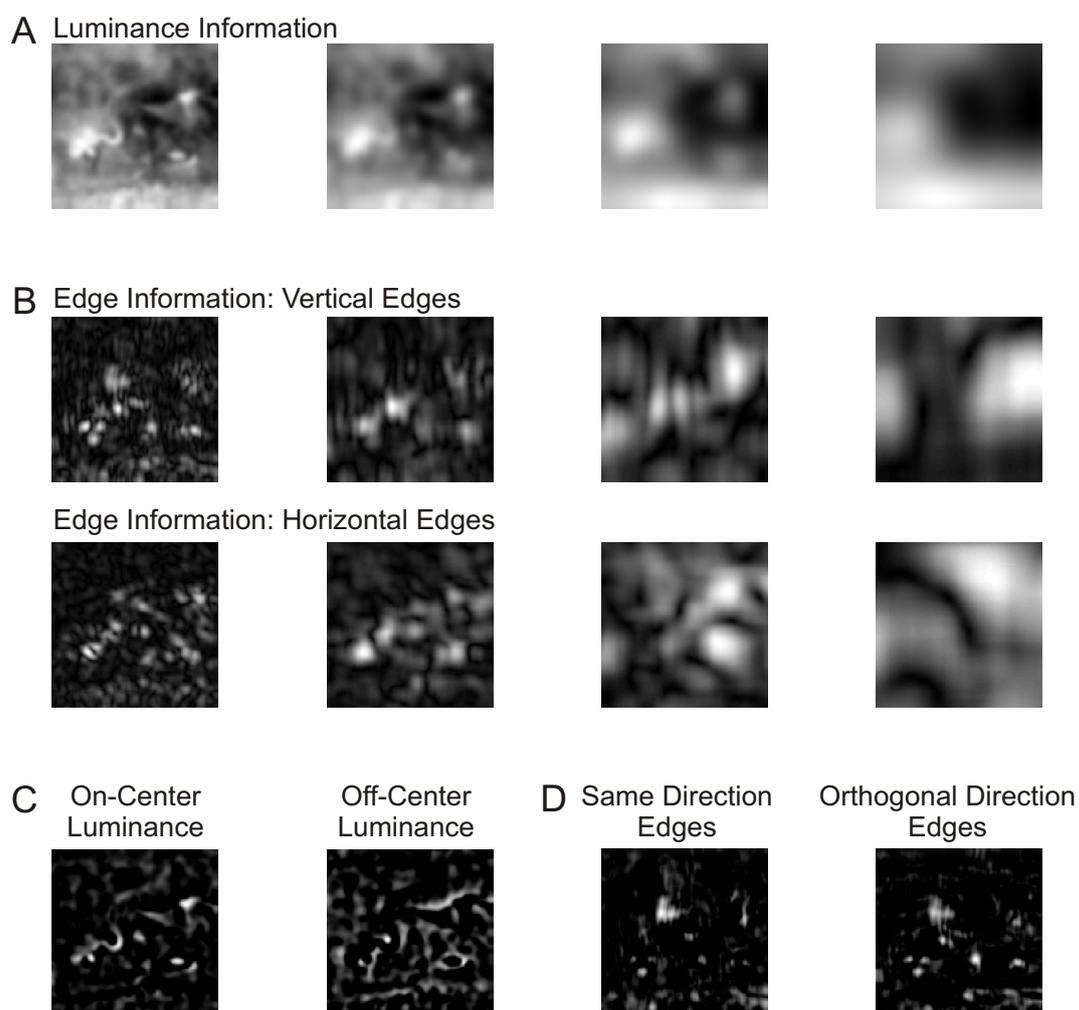


Figure 2.16: Exemplar analysis of a natural scene (scene N1). *A*, Luminance information computed at four spatial scales. *B*, Edge information at the same spatial scales. The upper row shows edges oriented vertically, the lower row horizontally oriented edges. *C* and *D*, Exemplar center-surround responses computed as the difference between scale 1 and scale 3 responses for luminance and edges.

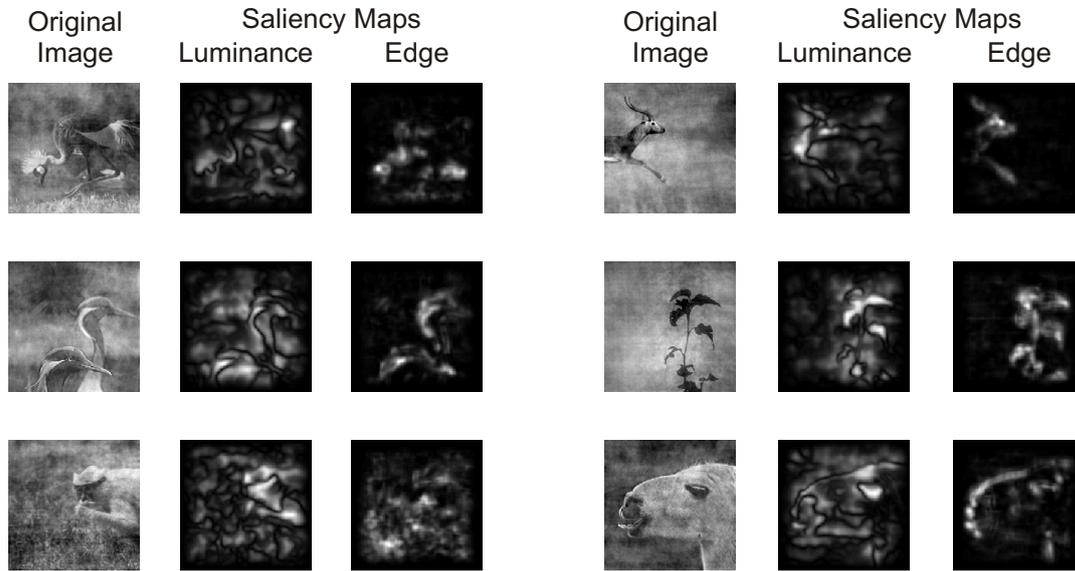


Figure 2.17: *Saliency maps for all natural scenes. Minima in the maps are shown in black, maxima in white.*

center-surround maps were generated for luminance. In the case of edge energy, two center-surround types were computed for each orientation. Here, either maps for the same orientation but at different scales were contrasted, or maps with orthogonal orientations and at different scales. Again, the difference maps were rectified, to create a total of 48 edge center-surround maps. For an illustration of these maps, see Figure 2.16C and D.

In the final step, center-surround maps were combined to create a saliency map for luminance and edges. The purpose of the saliency map is to represent the conspicuity or saliency of a visual feature at every location in the visual field by a scalar quantity. Since a large number of maps needs to be combined to generate saliency maps, simple summation of all individual maps seems problematic. A salient object may only elicit a strong peak of activity in one or a few feature maps, tuned to the features of that object. A linear combination of maps may then lead to masking of the salient object because of noise or less salient objects in other maps. Center-surround maps were therefore normalized before combining them to form the saliency maps. The normalization process was selected to promote maps with a small number of strong peaks of activity, while globally suppressing maps with numerous comparable peak responses. Each map was first scaled to the range of 0 to 1. The map's global and local maxima were next identified, and the average \bar{m} of all local maxima, excluding

the global maximum, was computed. Maps were then divided by $(1 - \bar{m})^2$. In the case of the luminance saliency map, each of the 12 center-surround maps was normalized individually; normalized maps were then added. For the edge saliency map, again all 48 center-surround maps were first individually normalized. The 12 maps belonging to one orientation were then summed, and the resulting four maps were again normalized before adding them to form the saliency map. The saliency maps computed for each of the natural scenes are shown in Figure 2.17.

2.5.5 Analysis of neurophysiology data

Single cell responses

During physiology sessions, the multi-unit activity was continuously registered at each of the electrodes. The multi-unit activity consists of two components: Superimposed on a background of low amplitude are large action potentials. The background reflects the activity of non-isolated neurons in the vicinity of the electrode tip, while the action potentials are generated from isolated neurons very close to the electrode. The activity of individual neurons therefore needs to be separated from the background activity. Also, multiple neurons can be close enough to the electrode to generate spikes larger than the background activity. The individual contributions of these neurons thus have to be identified before the rest of the analyses could be performed. This so called spike sorting was done for each electrode using a commercial software (Offline Sorter, www.plexoninc.com). First, spike waveforms were extracted from the continuous recordings by determining the time points at which the recorded signal amplitude exceeded a certain threshold. This separated spikes from the background activity. The rest of the spike sorting then served to separate the spikes from multiple neurons. A fixed time window was set around each threshold crossing, and only the data segments in these windows were kept for further analysis. The shape of all extracted waveforms was then characterized by a number of parameters like maximal and minimal amplitude or more complex shape descriptors. The shape of a waveform recorded from a neuron depends on the spatial arrangement of the neuron and the electrode tip. Hence, different neurons generate different waveforms. This in turn results in very similar shape parameters for the waveforms generated by one neuron, while the shape parameters for different neurons are more dissimilar. Similarities between waveforms can be assessed by plotting the values of all waveforms along two selected shape parameters. In these 2-D projections, waveforms of the same neuron build a large cluster. Depending on the selected projection, clusters generated from different units can be separated.

The projections resulting in the largest separation between clusters from different units was identified by inspection, and the clusters were outlined manually. All waveforms within a cluster were assigned to the same neuron. As end result, the program returned the time stamps for the spikes generated by each of the identified neurons.

All further analysis was done using Matlab. First, spike time stamps were converted into a rate code for each unit. The baseline firing rate and the stimulus driven firing rate were computed by counting the number of spikes in the 200 ms interval preceding stimulus onset, and in the interval from 100 to 400 ms after stimulus onset, respectively. Whenever net firing rates were analyzed, these were generated by subtracting the baseline activity from the stimulus driven firing rate. In addition, spike density functions were computed for each unit. These functions were computed by convolving the spike train of each trial with a Gaussian kernel:

$$p(t) = \sum_{i=0}^{N-1} f(i) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(t-i)^2}{\sigma^2}\right),$$

where t denotes the time relative to stimulus onset, and $f(i)$ is the spike train represented with a 1-ms resolution. $f(i)$ is equal to 1 if a spike occurred in the i th 1-ms interval, and 0 otherwise. N gives the total number of spike train bins. A standard deviation σ of 5 ms was chosen. The convolved signal p was averaged over all trials of the same condition. For each image, the spike density functions computed in different conditions were normalized by dividing them by the maximal value observed across all conditions for this image.

The spatial dependency of a neural response was characterized by computing a so called response image. The response image pooled all occluded conditions of one image. To compute the response image for a particular image, the maximal firing rate of a neuron was first determined across all conditions of all images. Then, the net response in each occluded condition of the selected image was compared against the maximal firing rate. If a condition evoked a net firing rate that was at least as large as 70% of the maximal response, the condition was considered further. Mask values in each condition normally varied between 0 and 1. For the following analysis, all mask values greater than 0 were set to 1, so that mask values were binary. Based on the binary masks, the response image RI was computed as

$$RI = \sum_i M_i.$$

The sum was carried out over the binary masks M from effective conditions. The response image was also transformed into a binary mask by setting all values larger than 0 to 1. A response image thus shows the image regions which were visible in at least one of the effective conditions.

The selectivity of a neuron was quantified in terms of the mutual information that the firing rate conveyed about which image had been presented in a trial, and by assessing the neuron's tuning curve. Both computations were performed separately for each condition. Mutual information was computed by first determining the number of spikes evoked in a selected interval after stimulus onset. In most cases, the interval from 100 to 400 ms after stimulus onset was chosen. The distribution of spike counts was then computed by combining the spike counts elicited by the four images, and binning the responses into 12 bins between the minimum and the maximum spike count. Bins were placed so that the first bin was centered on the minimum, and the last bin was centered on the maximum. From this response distribution, the mutual information could be derived as

$$I = \sum_{s=1}^S \sum_{r=1}^R p(r, s) \log_2 \frac{p(r, s)}{p(r)p(s)}.$$

R and S indicate the total number of response bins and images, respectively. $p(s)$ denotes the probability of occurrence of one of the images. Similarly, $p(r)$ is the overall probability of a spike count falling into one of the response bins. The joint probability of the occurrence of one of the images and a spike count falling into a specific response bin is given by $p(r, s)$. Since only a limited sample size is available to compute any probability, a systematic error is introduced in the computation. Different methods have been proposed to correct for this error. Here, the procedure suggested by Panzeri & Treves (1996) was implemented, and the following bias was subtracted from the computed mutual information:

$$B = \frac{1}{2N \log 2} \left(\sum_s \tilde{R}_s - \tilde{R} - (S - 1) \right),$$

where \tilde{R}_s denotes the number of “relevant” response bins for the trials with image s , i.e. the response bins in which the occupancy probability $p(r, s)$ is non-zero. In the same way, \tilde{R} denotes the number of response bins where $p(r)$ is non-zero.

To compute the tuning curves, images were rank ordered for each neuron according to the net firing rate they evoked in the unoccluded condition. This ordering was kept for the occluded conditions, and net firing rates were averaged across neurons as a function of stimulus rank. Rank ordering was also used to compare stimulus

preferences across conditions. Images were rank ordered according to their firing rates in the unoccluded condition, as well as in each occluded conditions. The best stimulus was then identified in the unoccluded condition. The rank of this best image could then be determined for the occluded condition.

For each neuron, the effectiveness of manipulating either the visible stimulus size or the diagnosticity were assessed. Each factor was treated independently. Effectiveness was quantified as the variability in firing rate explained by changes in one factor. For this analysis, net firing rates were computed for each trial of a particular image. Subsequently, trials were grouped according to the levels of the factor of interest, i.e. either according to whether they were diagnostic, or according to the visible stimulus size. The means of firing rates within each group were computed, as well as their variability:

$$V_{Group} = \sum_l \left(\left(\sum_i f_{il} \right) - \bar{f} \right)^2,$$

where f_{il} indicates the firing rate in the i -th trial of factor level l , and \bar{f} is the mean firing rate across all trials. The total variability present in the data is as usually given by

$$V_{total} = \sum_l \sum_i (f_{il} - \bar{f})^2.$$

Finally, the variance explained by one of the factors amounts to $V_{Group}/V_{total} \cdot 100\%$ (Bortz, 1993).

A neuron's selectivity could also be expressed in terms of an explained variance value. Here, trials were grouped according to which image was presented, taking only the unmasked images into account. Selectivity was then given as the variance explained because of the presentation of different images.

To characterize the timing of the neural response, the onset latency of each neuron was computed. Latencies were derived separately for each image and condition. To determine the response latencies, neural responses were characterized as peristimulus time histograms (PSTHs) with a resolution of 10 ms, computed for the time interval from 200 ms before stimulus onset to 500 ms after stimulus onset. Each PSTH bin gives the probability of a spike occurring within a particular 10 ms time interval. Baseline activity of a neuron was quantified by the mean and standard deviation of the PSTH values in the 200 ms prior to stimulus onset. The latency of a neuron was determined as the time of the first response bin exceeding the baseline mean plus 1.96 times the baseline standard deviation, followed by at least one more bin

fulfilling the same criterion. The chosen criterion implied that PSTH levels had to be statistically significant from the baseline level at $p < .05$.

Local field potentials

LFP signals were recorded with a sampling rate of 4.46 kHz. In a first preprocessing step, the signal was downsampled to 1000 Hz. Second, a low-pass filter was applied to remove slow drifts and other artifacts. Signals were filtered with a first-order Butterworth filter generating a band-pass between 5 and 80 Hz. This filtering provided more stable signals. However, the same results were obtained without additional filtering. Since different channels were often recorded with different gains, the data were normalized next. For this purpose, the mean and standard deviation of the LFP was determined during a 100 ms baseline interval preceding stimulus onset. The whole LFP data in a trial were then transformed by computing

$$Z(t) = \frac{L(t) - \mu_{Base}}{\sigma_{Base}},$$

where $L(t)$ was the original LFP time series, and μ_{Base} and σ_{Base} the mean and standard deviation during the baseline period, respectively. Each trial was normalized independently. As a result, the transformed LFP was centered around zero in the baseline periods, and specified amplitudes in units of standard deviation. When computing visual evoked potentials (VEPs), the transformed LFPs were averaged stimulus locked, i.e. by aligning different trials to stimulus onset before averaging.

The further analysis was restricted to three 20 ms time intervals during stimulus presentation. These time intervals were centered around the location of three prominent peaks present in the VEPs averaged across all images and recording sites. Referring to their polarity and latency, the three components were labeled N100, P130, and N200. To account for latency differences, the time intervals were set independently for each condition and monkey, with the restriction that for a monkey, the time windows were identical for diagnostic and non-diagnostic conditions of the same visible stimulus size. Table 2.1 lists the peak latencies for all conditions and both monkeys.

To analyze the LFP behavior in these time intervals, the VEP amplitudes in each trial were averaged across the 20 ms time periods. Based on these averages, the mean VEP amplitude across recording sites could be given for each of the time windows. In addition, the influence of the two stimulus manipulations on the LFP was quantified in terms of the explained variance. In principle, the same analysis was carried out as for the single units. Instead of the firing rates, the average LFP during one of the 20 ms epochs was subjected to the analysis. The average LFP

Table 2.1: Latency of the three identified VEP peaks, given in ms after stimulus onset. Diagnostic and non-diagnostic conditions of the same size have been collapsed. Conditions are labeled by how much of the stimulus was visible.

Condition	G00			B98		
	N100	P130	N200	N100	P130	N200
Full	106	130	195	127	161	206
10%	133	166	226	141	177	238
30%	119	142	198	131	164	219
50%	106	135	197	129	165	216

amplitude was computed in each trial for the three time windows. Based on this data, the variances explained by diagnosticity and visible stimulus size could be computed at each site.

Chapter 3

Results

3.1 Behavioral results

The behavioral experiments tested the influences of partial occlusion on the behavior of monkeys who had to identify members of various stimulus sets. Both simple, geometric stimulus sets, as well as complex natural scenes were tested. The consequences of partial occlusion on the discrimination performance were used to identify sources of information in the stimuli. Results obtained for monkey observers were compared with the behavior of human observers; in addition, diagnostic regions for a task were contrasted with regions of interest evident from visual inspection of the images.

3.1.1 Effects of occlusion on the discrimination of simple geometric shapes

The first experiment was designed to test the Bubbles paradigm with simple stimuli before using natural scenes. It therefore used a set of simple geometric shapes (see Figure 2.3A in Section 2.2.1). Shapes were constructed so that each of them had a single diagnostic feature. Shapes could only be identified by the position of the small rectangle. As a consequence, when an observer has to discriminate between partially occluded versions of these shapes, occlusion of the small rectangle will render the shape unidentifiable. Occlusion of other shape regions should have no systematic effect. This precise control of diagnostic stimulus regions allowed an easier interpretation of the results, in contrast to the complex outcomes possible

with natural scenes. The results could therefore be used to test whether the Bubbles paradigm works in monkeys.

The monkeys were initially trained to discriminate the three shapes. In the discrimination task, shapes were presented for 300 ms. This is the same timing as used by Kovács et al. (1995). After stimulus presentation, the monkeys had to execute a saccade to one of three targets displayed in the periphery. Each target was associated with one of the shapes. The monkeys thus indicated which stimulus they perceived by selecting a particular target. Fixation of the correct target resulted in the delivery of a drop of juice as reward. When the monkeys were proficient in discriminating the stimuli, they were transferred to the Bubbles task. Stimuli were now presented behind trial-unique masks partially occluding the stimuli. The monkeys continued to discriminate between the masked shapes. As before, stimuli were presented for 300 ms, followed by the three response targets. Again, the monkeys had to make a saccade to the correct response target to receive a reward.

As in the original paradigm developed by Gosselin & Schyns (2001), the trial-unique masks consisted of an occluding surface, punctured by round windows (“bubbles”) of a fixed size through which the underlying stimulus could be seen. Mathematically, masks were images with the same size as the original images. Each mask pixel had a value between 0 and 1. To generate a partially occluded image, the gray-scale values of the original image were mixed with a gray background. Mask values determined how much the image values contributed to the mixture. Mask values of 0 corresponded to completely covered image regions, while mask values of 1 indicated completely visible regions.

Bubbles were randomly positioned on every trial. Their number was adjusted such that the visible stimulus portion was large enough to allow a correct stimulus identification in 70% of the trials. Bubble diameter was set to 1.35 deg, which corresponded to about 23% of the stimulus size. Masked stimulus presentations were interleaved with unmasked ones to assure that the monkeys maintained their performance level in the basic task. Consequently, errors in masked trials were mostly due to the influences of the masks.

In each masked trial, it was recorded which stimulus had been shown, which mask had been used, and whether the monkey had identified the stimulus correctly or not. After collection of sufficient data, trials were first sorted according to stimulus identity. The trials for each stimulus were then further divided into correct and incorrect trials depending on the monkey’s response. Finally, masks in correct trials were compared to masks in incorrect trials to identify image regions where occlusion

had a systematic effect on performance. Because all masks had the same size, analyses could be performed for each pixel separately.

First, the average over all masks from correct trials was computed, as well as the average over all masks from incorrect trials. The two averages allowed a quantitative description of the effects of the masks. They indicated which image regions were consistently covered or visible in the respective trial types. Second, the distribution of mask values at a pixel was used to quantify differences between correct and incorrect trials. This analysis was based on the following reasoning: If occlusion of an image pixel has a systematic effect on performance, this image pixel will always be covered in incorrect trials, while it is visible in the correct trials. The mask values determine how much of a pixel is visible. At the hypothetical pixel, mask values in correct trials will thus be clustered around 1, while mask values in incorrect trials will be near 0. More generally, if occlusion of a pixel systematically affects the performance, mask values will have different distributions in the correct and in the incorrect trials. However, if the visibility of an image pixel has no effect on performance, similar mask values – and hence similar distributions – will be observed in correct and incorrect trials.

An analysis of the distribution of mask values showed this assumption to be justified. The box plots in Figure 3.1 summarize how the mask values were distributed at different image pixels of scene G1, analyzing the data collected for monkey B98. The averages over masks from correct and incorrect trials were taken as a qualitative description of the occluder effects. If these averages are very different at a particular pixel, the occlusion of the pixel presumably has robust effects on the behavior. Four pixels were selected based on the averages. Two pixels fell into image regions with different average values in correct and incorrect trials. The other two were selected in image regions without differences. Consistent with the above hypothesis, mask value distributions were different at those pixels at which the averages indicated occlusion effects. For the other pixels, very similar distributions were observed.

A quantitative assessment of the mask influences could therefore be based on a test that compared the mask value distributions between correct and incorrect trials. Mask values were restricted to the range from 0 to 1, and did not follow a particular distribution because they were simply a product of how bubbles were distributed across a mask. A non-parametric test seemed therefore appropriate, and the Kolmogorov-Smirnov test was selected for the analysis. This test gives the probability that two sample distributions could have been generated by drawing samples from the same population. It is sensitive to differences in any of the parameters affecting a distribution's shape (Siegel, 2001). Since a test was computed at each pixel, the resulting p-values had to be Bonferoni corrected for the number

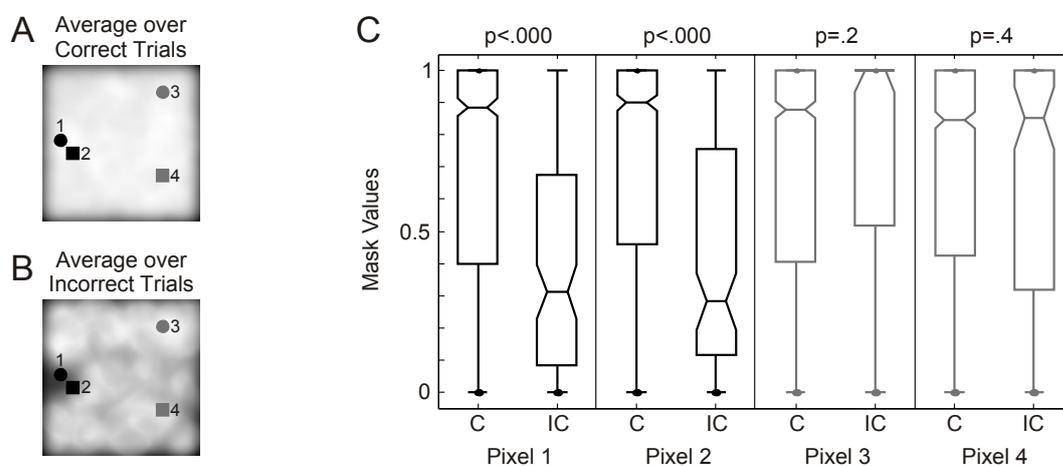


Figure 3.1: The distribution of mask values at different image locations (data from scene G1, experiment with monkey B98). A, Average mask computed from correct trials. B, Average mask computed from incorrect trials. Four image pixels (denoted as Pixel 1 - 4) were selected. Pixel 1 and 2 fall on image regions where the averages show differences between masks in correct and incorrect trials, while at Pixel 3 and 4, the masks do not seem to influence the performance. C, Box plots of the mask value distribution for each of the pixels. C indicates correct, IC incorrect trials. The box plots show lower quartile, median, and upper quartile by horizontal lines; whiskers extend 1.5 times the interquartile range. p -values rate the difference between the distributions (computed by a Kolmogorov-Smirnov test).

of comparisons. The corrected p-values were then used to identify image regions where occlusion exerted a systematic effect on performance. The obtained results were not an artefact of the selected test. To verify this, the Kolmogorov-Smirnov test was replaced by the Mann-Whitney U test. The relationship between occlusion and behavior remained similar in both cases, i.e. the same regions were identified to show a systematic or non-systematic effect of occlusion.

Figure 3.2 shows how occlusion affected the discrimination performance of monkey B98. These results were based on a total of 1400 masked trials per image, of which between 10 and 20% were incorrect. In the intermingled unmasked trials, the error rate was between 2 and 4%, confirming that the monkey was performing well in the basic task. Error trials could hence mainly be attributed to the influence of the mask. Little structure was visible in the average masks from correct trials. However, for each shape the average mask from incorrect trials contained a dark patch over the position of the small rectangle. This indicates that in trials in which the monkey failed to identify a shape, the small rectangle was occluded. The quantitative analysis confirmed this observation. Statistical maps showed focal regions with systematic differences in masks from correct and incorrect trials. These regions were located at the position of the small rectangles. To interpret the statistical results more easily, p-value maps were summarized by identifying shape regions for which the statistical comparison yielded p-values smaller than .01. These diagnostic regions contained the small rectangle for all three shapes.

Monkey G00 was also tested with the geometric stimulus set. The error rates in the masked trials were slightly higher for G00 than for B98. 2100 masked trials were collected, with 20 to 30% of them being incorrect. Performance in the unmasked trials was again very good, and error rates were between 1 and 4%. Spatial influences of occlusion are shown in Figure 3.3. The results for shapes G1 and G2 confirmed the pattern observed in monkey B98: In both cases, the only shape region with a systematic effect of occlusion was the small rectangle. This was obvious both in the average masks from incorrect trials, as well as in the statistical comparison between masks from incorrect and correct trials.

A different pattern of results emerged for shape G3. Both the average mask from correct trials, and the average mask from incorrect trials were homogeneously gray, and no shape region could be identified where the average masks distinguished between trials with successful or unsuccessful identification. As expected, the statistical comparison also did not yield a significant result, suggesting that occlusion had no systematic effect on the identification of this shape. In conclusion, the Bubbles results suggested that monkey G00 identified shapes G1 and G2 by the position of the small rectangle, but used no such diagnostic feature to identify G3. Interestingly,

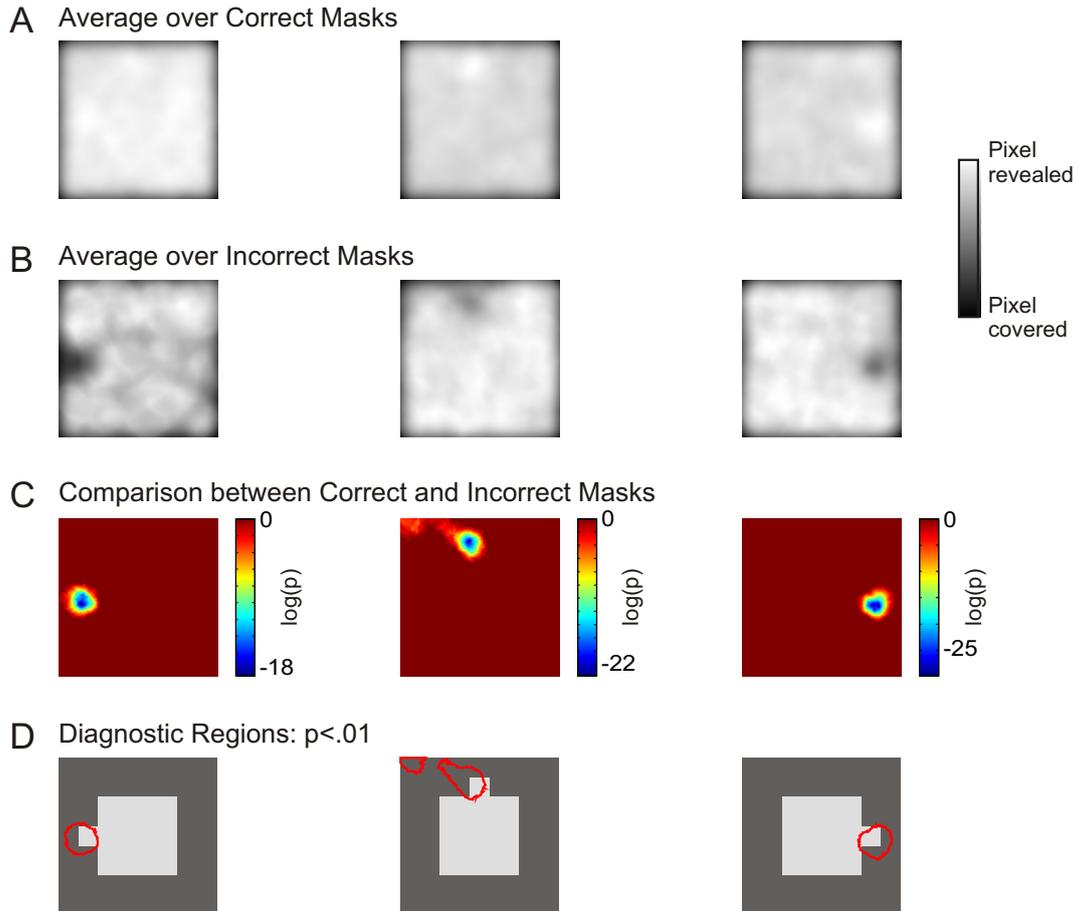


Figure 3.2: *Bubbles* results for monkey B98, discriminating a set of geometric shapes. Each column contains the results for the shape shown in D. A, Average over the masks from correct trials. B, Average over the masks from incorrect trials. Dark grays in A and B indicate that a pixel was covered, while bright grays show pixels that were not occluded by the average mask. C, Results of a comparison between the mask value distribution from correct and incorrect trials by means of a Kolmogorov-Smirnov test. Plotted is the logarithm of the Bonferoni corrected p-values. D, Diagnostic regions. Each line includes the pixels with Bonferoni corrected p-values smaller than .01.

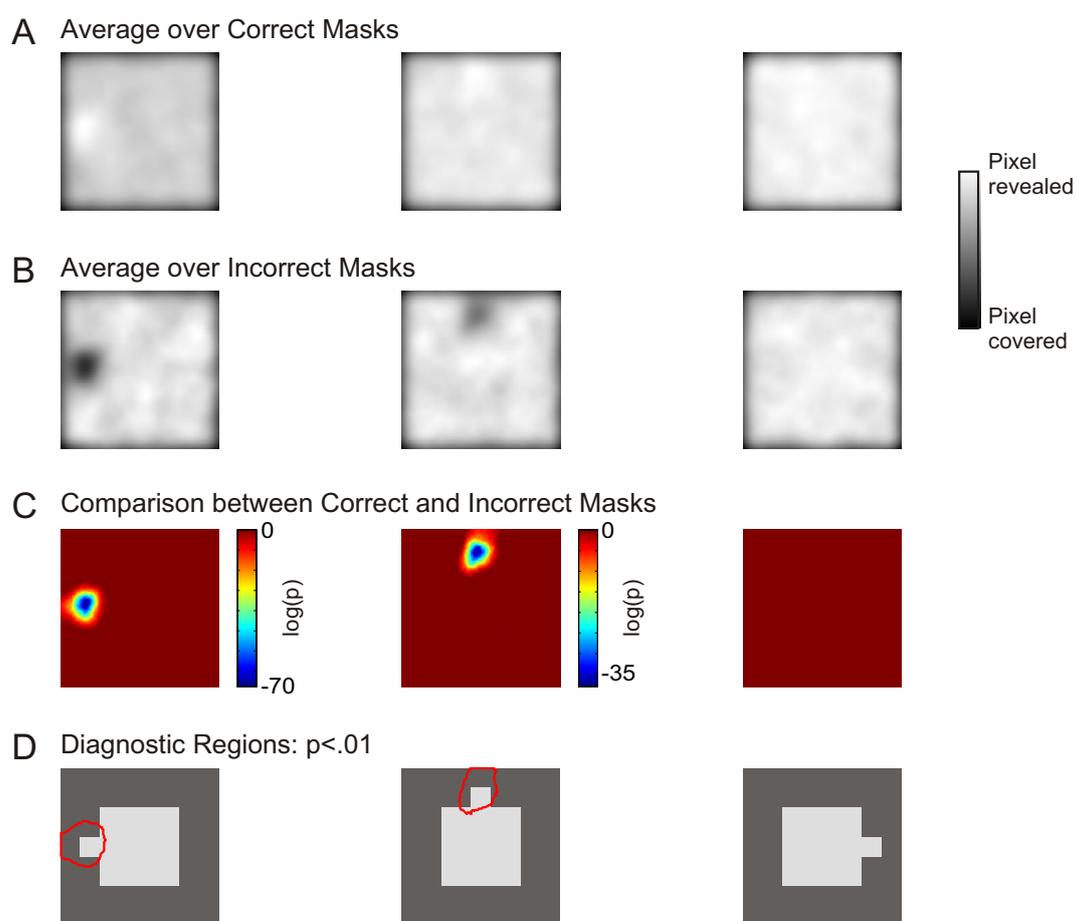


Figure 3.3: Influences of occlusion on the discrimination performance of monkey G00. Layout of the plot is identical to Figure 3.2.

this difference in the treatment of the stimuli was not accompanied by a difference in recognition rates. About 30% of the masked trials were incorrect for G3. This was the largest error score for the three shapes, but the error scores for the other shapes were not very different (21 and 24% for G1 and G2, respectively). Furthermore, in case of the unmasked trials G3 was identified with the smallest amount of errors (about 1% error trials).

There are two equally successful strategies to solve the discrimination task. In the first strategy, each shape is identified by its diagnostic feature. In the second strategy, only two of the shapes and their diagnostic features are remembered by the observer. The third shape is selected whenever a shape cannot be identified as either of the two remembered shapes. This latter strategy seems consistent with the observed influences of occlusion for monkey G00. It also matches the distribution of errors made during the task. Table 3.1 suggests that when the monkey failed to identify G1 or G2, these shapes were classified mostly as G3. This seemed to be the case for both masked and unmasked presentations. To verify the significance of these observations, the error trials were more closely analyzed. If the monkey made an error upon presentation of shape G1, he erroneously chose either shape G2 or G3. If the hypothesis is correct and shape G3 was encoded as “not G1 or G2”, then the number of error trials in which G3 was selected should be significantly higher than the number of trials with G2 responses. This was indeed the case (χ^2 -test for the errors on G1, masked: $\chi^2(1) = 8.82$, $p=.003$; errors on G1, unmasked: $\chi^2(1) = 18.69$, $p<.000$). Similarly, incorrect responses to G2 involved G3 significantly more often than G1 (errors on G2, masked: $\chi^2(1) = 29.77$, $p<.000$ errors on G2, unmasked: $\chi^2(1) = 5.00$, $p=.03$). Finally, misclassifications of G3 led to equal numbers of G1 and G2 responses (errors on G3, masked: $\chi^2(1) = 0.01$, $p=.90$; errors on G3, unmasked: $\chi^2(1) = 1.00$, $p=.32$). This shows that shapes that could not be identified as G1 or G2 were classified as G3. This pattern is consistent with a response scheme in which only G1 and G2 were actively encoded.

To confirm that this response model indeed is consistent with the observed Bubbles results, a model observer was constructed that implemented the same decision rules (see Figure 2.14 in Section 2.5.2 for an illustration of the model). Thus, the model identified G1 and G2 by their small rectangle, and selected G3 as response when a stimulus could not be identified as G1 or G2. The model observer used two templates to identify the shapes. One template was matched to shape G1, and contained a “detection window” at the position of the small square in G1. The second template was similarly matched to G2, and also contained an appropriately placed detection window. When a stimulus was presented to the model, it applied both templates to the stimulus. For each template, the value of a decision variable

Table 3.1: *Response matrix for monkey G00. The table shows with which percentage the monkey selected one of the three response alternatives after presentation of a particular stimulus.*

Resp. Stim.	Presented Stimulus					
	Unmasked Trials			Masked Trials		
	G1	G2	G3	G1	G2	G3
G1	96.5	0.6	2.9	79.3	8.9	11.9
G2	0.4	98.4	1.2	9.0	76.2	14.8
G3	0.8	0.5	98.7	14.7	14.9	70.4

was computed. The decision variable quantified the certainty of the model observer that a small rectangle was present in the detection window of the template. Each decision variable was based on the stimulus brightness in the respective detection window, plus a random noise term. The random noise term was necessary because real observers do not perform at a 100% correct even under optimal conditions. The random noise term was drawn from a Gaussian distribution with mean 0 and a fixed standard deviation. It could increase or decrease the value of the decision variable. The two decision variables were then used to generate a response to a presented stimulus. If the value of the decision variable for the template matching G1 was above a fixed threshold, and the value for the template matching G2 was below the threshold, the model identified a stimulus with G1. Similarly, G2 was detected as a value above threshold for the decision variable computed from the second template, and a value below the threshold for the first. Finally, G3 was chosen whenever both decision variables had values below the threshold.

The model was then confronted with partially occluded versions of the three shapes. Since the performance of the model was compared against the one of monkey G00, the exact same stimulus material was used. Thus, the same stimuli shown before to monkey G00 were now shown to the model observer. Two parameters needed to be chosen before the model could be used. These were the threshold for the decision variables, as well as the standard deviation for the distribution from which the random noise term was drawn. Both parameters affected the performance level of the model. The threshold was set such that without the noise term, the model identified a shape correctly if a quarter of its small square was visible. The noise term was then adjusted so that the model performed with the same overall error rates as monkey G00. The values for both parameters are given in Section 2.5.2. After every presentation of a masked stimulus, the model decided which stimulus had been presented. The rest of the analysis was then performed identically to the one for the monkey's data. Results are shown in Figure 3.4. As for the monkey,

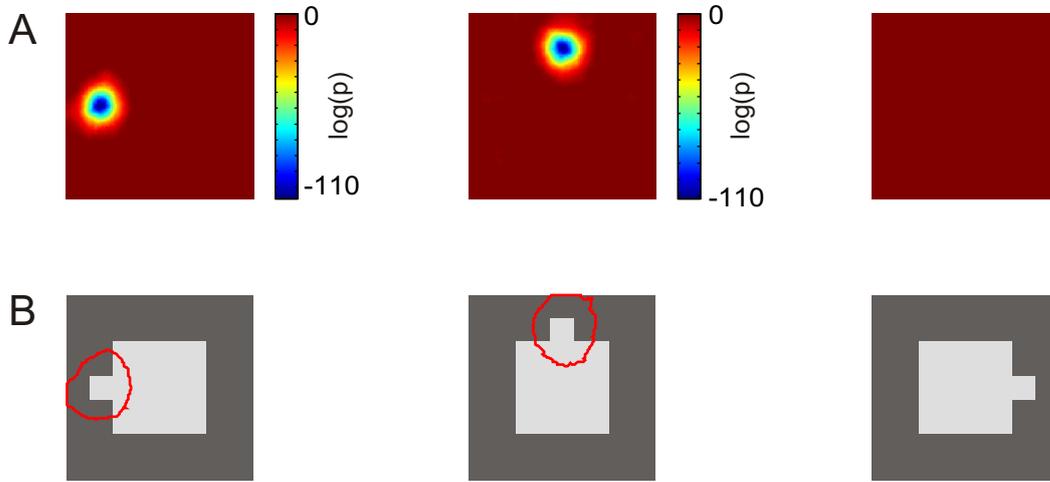


Figure 3.4: Results of the Bubbles testing for a model observer. *A*, p -values from a comparison between masks from correct and incorrect trials using the Kolmogorov-Smirnov test. *B*, Diagnostic regions. All p -values are Bonferroni corrected.

occlusion had systematic effects only for the first two shapes. Diagnostic regions included the two small rectangles in these cases. For G3, occlusion had no consistent influence on identification performance. These results confirm that a model that only identifies two of the shapes by their diagnostic features, and detects the third shape by excluding the other alternatives, is compatible with the Bubbles results for monkey G00.

In summary, the first experiment yielded the following results: For monkey B98, the effects of partial occlusion on the discrimination of simple geometric shapes were as predicted. In most cases, partial occlusion did not affect the discrimination performance. However, when the single diagnostic feature of a shape was occluded, the monkey reliably failed to identify the shape. For monkey G00, results were similar for two of the stimuli. These were identified by their diagnostic features. The third shape was recognized not because of its diagnostic feature, but by excluding the other two response alternatives. This different treatment of the shapes would not have been detected based on the error rates alone, especially in the case of unmasked presentations. To conclude, the results of both monkeys show that monkeys treat shape regions differently when performing a task, with some regions containing more information than others. At the same time, the experiments confirm that Bubbles can be used with monkeys to identify how information is distributed across a stimulus, revealing the monkeys' strategies.

3.1.2 Distribution of information across natural scenes for monkey observers

After the first experiment had established that the Bubbles paradigm could be used with monkey observers, it was used to determine the distribution of information across natural scenes. Two sets of natural scenes were used in the second experiment. The monkeys worked with one set at a time: They were first trained to identify the members of one set correctly, and then tested with the Bubbles paradigm, before training and testing them on the second set. The selected natural scenes were gray-scale pictures of an object (mostly animals and one plant; see Figure 2.3) in front of a non-uniform background. Images were normalized to have the same Fourier amplitude and to span the same range of gray-scale values. Nonetheless, images were very different locally, containing many possible sources of information. All parameters affecting the Bubbles paradigm were left identical to the first experiment.

The results for both sets are shown in Figure 3.5 for monkey G00, and in Figure 3.6 for monkey B98. For monkey G00, the comparison of masks from correct and incorrect trials yielded one region for each image in which occlusion robustly affected the task performance. The only possible exception is scene N6 in the second set, for which only a small region reached significance. Diagnostic regions always included the object present in the scene. For three of the the five animal pictures, the diagnostic region fell onto the head or the eye of the depicted animal.

The results obtained from monkey B98 were rather different. Again, for all but one of the scenes, a region in the image could be identified where occlusion systematically led to identification failures. However, with the exception of image N5 in the second set, diagnostic regions did not include or only partially included the object shown in a scene. Instead, diagnostic regions seemed to include scene parts with either very distinct luminance (like the dark patches in scenes N2 and N7) or contrast edges (as in scene N3). Monkey B98 therefore used a completely different strategy, which at first glance seems very unexpected. However, since images differed in many ways in their low-level properties, using these properties is an equally successful strategy to identify an image as using the scene's main object.

The particular strategy followed by monkey B98 seemed not to be a consequence of a learning process that took place during the Bubbles testing. Instead, an analysis of the data collected during the first session showed that the monkey already used information from the same image regions. In this first session, about 130 trials were collected per image, of which about 30 were incorrect. The sample size for the incorrect trials is too small to allow a statistical comparison. The analysis was

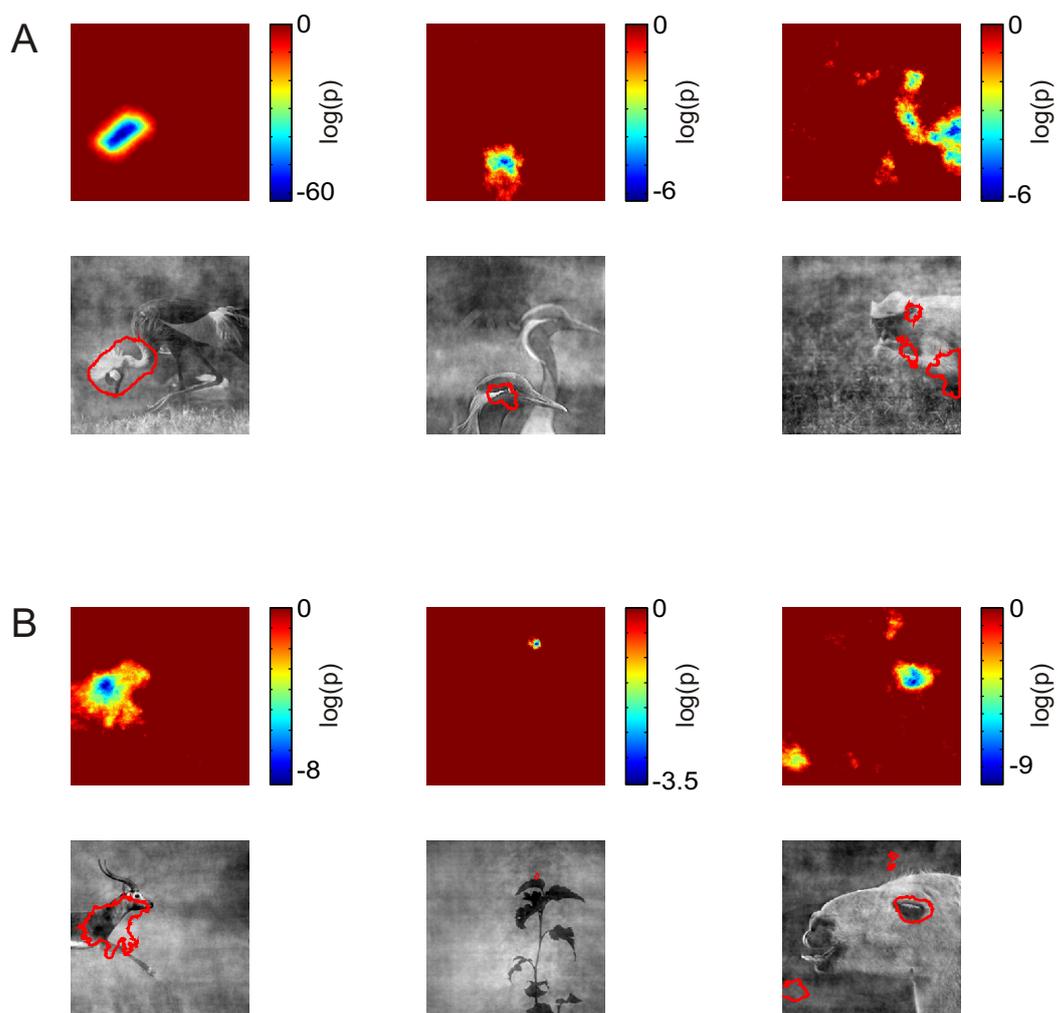


Figure 3.5: *Effects of occlusion on the discrimination of natural scenes, monkey G00. A, Natural scene set 1. B, Natural scene set 2. The upper row in A and B plots the logarithm of the p -values obtained from the Kolmogorov-Smirnov test. In the lower row, regions with p -values below 0.01 are encircled on the natural scenes. As before, all p -values are Bonferroni corrected.*

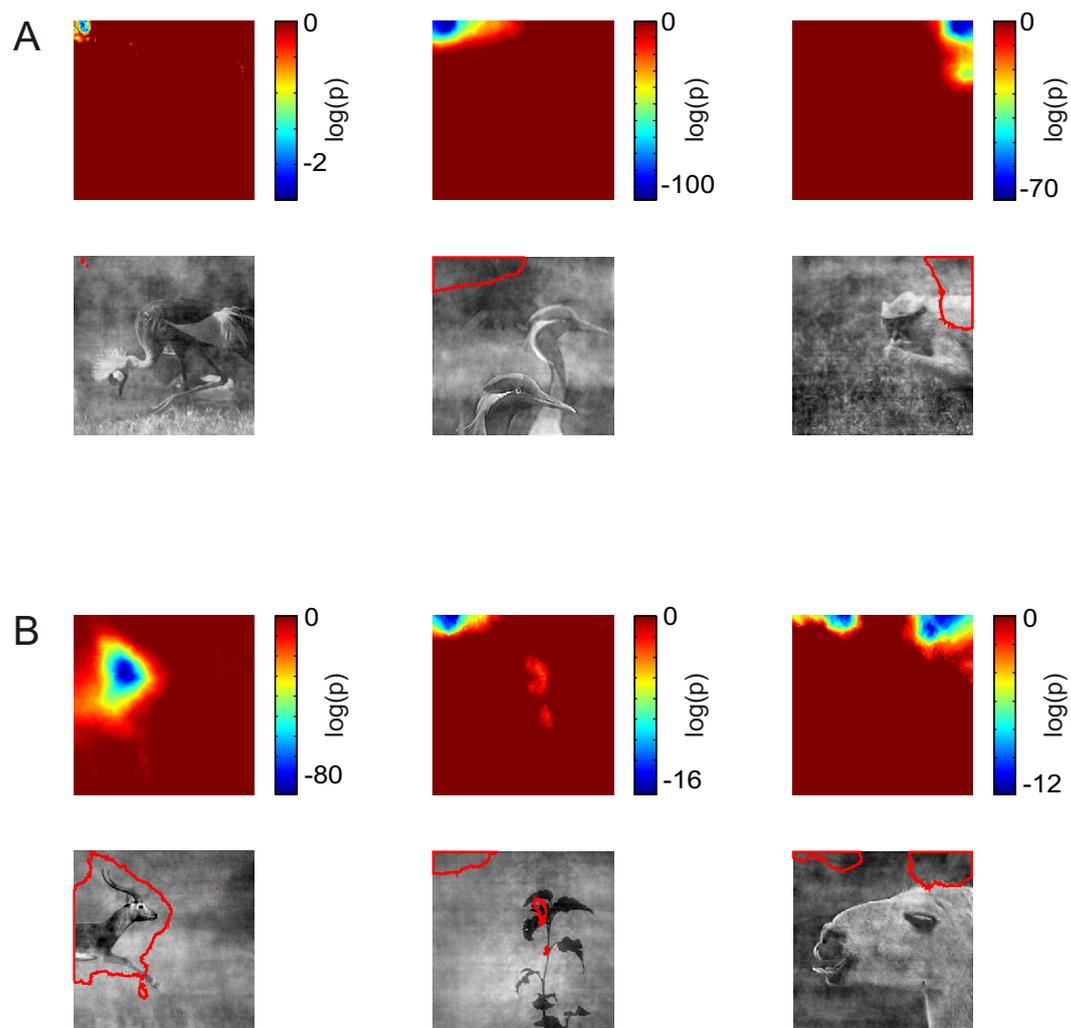


Figure 3.6: *Distribution of information across the natural scenes for monkey B98. A, Natural scene set 1. B, Natural scene set 2. The upper row of each panel shows the results of the statistical test (Bonferroni corrected p-values), the lower row the diagnostic regions.*

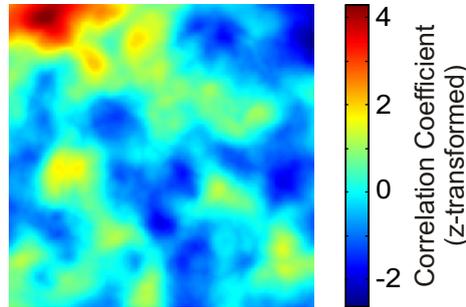


Figure 3.7: *Initial distribution of diagnosticity for scene N2. The image shows the z-transformed correlation coefficient between a mask value for a pixel, and the correctness of the monkey’s response. The data come from the first testing session with monkey B98. High correlation coefficients show that pixels were visible in correct, and occluded in incorrect trials.*

therefore restricted to computing the correlation between the mask value at a given pixel, and the correctness of the observer’s response. As an example, the results from the first testing session are shown in Figure 3.7 for scene N2. Image regions with a high correlation coefficient correspond to regions which were occluded during incorrect trials, but visible during correct ones. As can be seen, the upper left corner was the image region attaining the highest correlation coefficients. Identical to the statistical results obtained from many days of testing, these results indicate that occlusion of the left corner disrupted performance. The monkey’s decision strategy was therefore already evident on the first day of testing.

For the first natural scene set, monkey B98’s strategy was also not generated because this set was the first that the monkey learned to discriminate. After testing had been completed with the second set of scenes, monkey B98 was re-tested on the first set. The results of the re-test are given in Figure 3.8. A comparison of Figure 3.6 and 3.8 shows test and re-test to be in good agreement. Considering only scenes N2 and N3, 84% of a diagnostic region of the re-test was also diagnostic in the first testing. Since several months lay between test and re-test, the results suggested that the distribution of information across the natural scene was very stable for the monkey.

So far, the effects of occlusion have been described by determining the image regions where the statistical tests showed masks to have a reliable effect on performance. These regions have been taken to be the diagnostic image regions, i.e. the image regions from which an observer draws information to identify an image. Judging

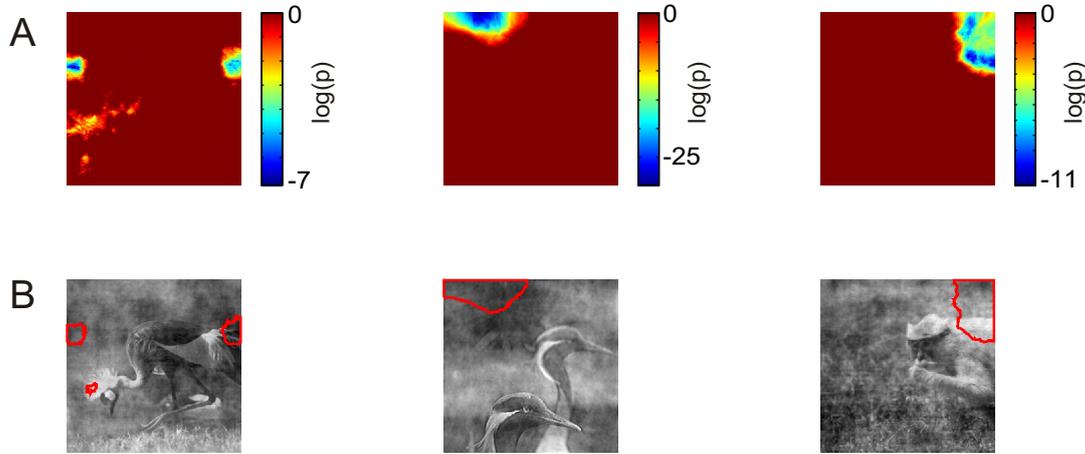


Figure 3.8: *Re-test of monkey B98 on natural scene set 1. As before, the upper row depicts the logarithm of the p-values computed with a Kolmogorov-Smirnov test, while the lower row shows diagnostic image regions.*

from the average masks from correct and incorrect trials, occlusion introduced differences in performance because diagnostic regions were occluded in incorrect trials, and visible in correct ones. However, during the actual testing images were shown behind masks with randomly placed bubbles. This random placement makes the number of trials very low in which masks did exactly cover or uncover all portions of the diagnostic regions. It therefore seemed necessary to confirm the conclusions drawn from the Bubbles experiments. For this purpose, masks were constructed based on the Bubbles results. These masks either covered or uncovered the diagnostic regions. The discrimination performance of the monkeys was then tested with the constructed masks. If diagnostic regions are indeed meaningful image portions to the monkey, a stimulus in which the diagnostic region is visible, while the rest of the image is covered (“diagnostic conditions”), should be identifiable. In contrast, covering the diagnostic region and uncovering some other portion of the image has to result in a stimulus that the monkey cannot identify correctly (“non-diagnostic conditions”).

To construct these masks, a number of parameters had to be known. One was the placement of the occluder. This was given by the Bubbles results, since masks were designed to cover either diagnostic or non-diagnostic image regions. The shape of the mask similarly followed from the Bubbles results. However, the appropriate size of the occluder was not as obvious. One possibility was to construct masks that covered all image regions where the p-value was below .01, as has been used to

outline diagnostic regions so far. However, this would generate masks of different sizes for different images. A different method was therefore chosen to determine the extent of the occluders. Occluders were here constructed so that they revealed a fixed amount of each image. To test the influence of the occluder size in addition to the influence of the occluder placement, three sizes were chosen. Occluders revealed either 10, 30, or 50% of an image.

Considering all requirements on the occluder generation, the following method was chosen to derive them: As before, mask computation was based on the p-values computed from the Kolmogorov-Smirnov test. Any mask was initially set to completely cover the whole image. To generate a mask that revealed the diagnostic image portions, the mask was made transparent over image regions where the corresponding p-values were very low. Three different mask sizes were generated by selecting the pixels with p-values belonging to the lowest 10, 30, or 50% of all p-values to be visible. The constructed masks consequently covered 90, 70, or 50% of an image. Similarly, masks for the non-diagnostic conditions were built by selecting image pixels with p-values belonging to the highest 10, 30, or 50%, and making the masks transparent at these locations. By these means, masks were generated that were non-overlapping in the two conditions, differed in the behavioral relevance of the visible image portion, but revealed the same amount of the underlying image.

Because of the different occluder placements, the generated stimuli differed in terms of their low-level characteristics like overall contrast or luminance. These differences were removed by equalizing the luminance, as well as the overall contrast, of the visible image regions across all stimuli. In summary, seven conditions were constructed for each image: Images could either be shown unoccluded, behind one of three masks that revealed diagnostic image parts, or behind one of three masks revealing non-diagnostic regions. The three masks of each group uncovered 10, 30, or 50% of the image. Since the masks were computed from the Bubbles results, each monkey had its own specific stimulus set. Figure 2.9 and Figure 2.10 show some of the generated stimuli for both monkeys.

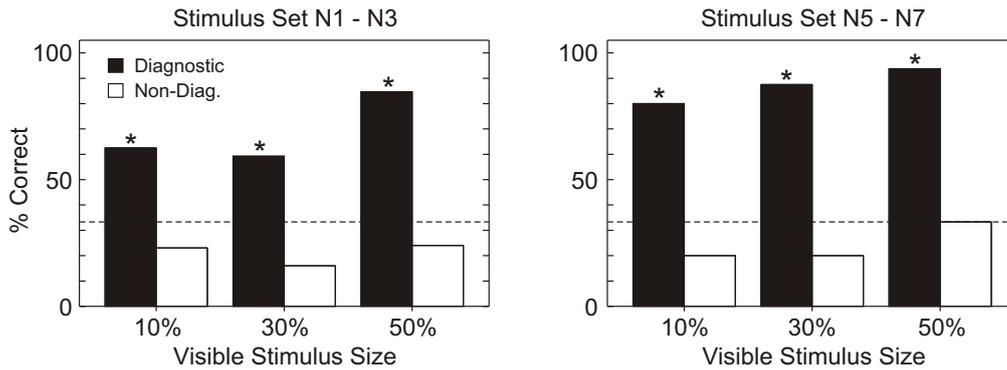
As the monkeys had already been trained to discriminate partially occluded images, their discrimination performance on the constructed stimuli could be tested by simply replacing the standard Bubbles masks with the constructed masks. The rest of the Bubbles paradigm remained mostly unchanged. Two minor changes were required because of the nature of the generated masks. While the normal Bubbles masks were controlled such that correct responses were made in 70% of the trials, masks were now built so that half of them explicitly interfered with the recognition performance (the non-diagnostic conditions). Usually, correct responses were rewarded for masked and unmasked trials, and incorrect responses were indicated

to the monkey by a sound. Applying the same reward scheme here would result in an overall lower reward probability, which poses problems because of the monkeys' motivation to perform the task. On the other hand, indicating incorrect responses may help the monkeys to eventually form an association between the non-diagnostic conditions and the correct response, thereby distorting the results. The reward scheme was therefore changed for the masked trials only; the monkey received the usual feedback on unmasked trials. On masked trials, 50% of the trials were now randomly rewarded irrespective of the correctness of the response. For the same reasons, a second change was introduced in the paradigm, and masked stimuli were now shown less often than unmasked stimuli. Only few repetitions were run for each masked stimulus, so that the monkeys would not change their response pattern to these stimuli.

The monkeys were tested with both stimulus sets. Since no diagnostic region could be determined for image N1 in the case of monkey B98, this image was excluded from the analysis of the performance of monkey B98. Error rates were computed for each mask size. As only few repetitions were run for each stimulus, all stimuli were considered together to increase the sample size. Figure 3.9 shows the error rates for diagnostic and non-diagnostic conditions as a function of mask size. For each size and condition, the error rate was tested against the chance level of 33% correct responses using a χ^2 -test. For monkey G00, both stimulus sets showed the same pattern: Stimuli in which the diagnostic regions were visible, were identified correctly significantly more often than chance ($\chi^2(1) \geq 8.17$, $p < .000$ for all sizes in both stimulus sets). This was the case even for the smallest mask size. In contrast, masking the diagnostic conditions resulted in stimuli that were not identified any better than chance ($\chi^2(1) \leq 3.38$, $p > .06$ for all sizes in both stimulus sets). These results showed a clear distinction between diagnostic and other image regions in importance for the monkey's task performance.

The same distinction between diagnostic and non-diagnostic conditions was obtained for monkey B98 when tested with the second stimulus set containing images N5 - N7: Again, the performance on diagnostic conditions was significantly better than chance, irrespective of mask size ($\chi^2(1) \geq 32.82$, $p < .000$ for all three mask sizes). For the non-diagnostic conditions, performance with the smallest mask size was worse than chance ($\chi^2(1) = 5.82$, $p = .02$); all other non-diagnostic conditions were identified at chance level ($\chi^2(1) \leq 1.69$, $p \geq .2$ for the two conditions). For the other stimulus set, diagnostic versions of scenes N2 and N3 were again identified better than chance at all three sizes ($\chi^2(1) \geq 14.24$, $p < .000$ for all conditions). This was the case even though the diagnostic regions did not contain the scene's object. For these scenes only the smallest non-diagnostic condition was efficient in hindering

A Performance of Monkey G00



B Performance of Monkey B98

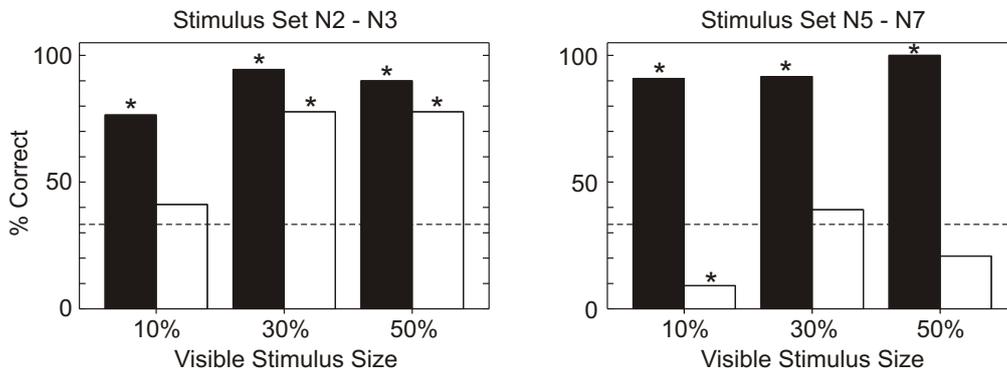


Figure 3.9: Verification of the Bubbles results. Stimuli were constructed in which the diagnostic regions were either visible (diagnostic condition, black bars), or invisible (non-diagnostic condition, white bars). Different amounts of the images were visible through the masks. Bars show the percentage of correct responses, collapsed across all images in a set. The dashed line shows the chance level, and stars indicate conditions in which the response rate deviated significantly from chance ($p \leq .05$). A, Results for monkey G00 on the two stimulus sets. B, Results for monkey B98. Image N1 was excluded from the analysis for this monkey.

stimulus identification ($\chi^2(1) = 0.47$, $p=.49$). When non-diagnostic masks revealed 30 or 50% of the stimuli, the monkey became significantly better than chance in identifying the stimuli ($\chi^2(1) = 16$, $p<.000$ for both conditions). This suggested that task relevant information was distributed very broadly across the scenes; yet the effects at the smallest stimulus size confirmed a special role for the diagnostic region.

To summarize, regions in natural scenes could be identified using Bubbles where occlusion systematically affected the monkeys' discrimination performance. These regions were unique for each monkey, and suggested a different involvement of low-level information in the task performance. Stimuli in which the diagnostic regions were covered were not identifiable by the monkeys. Occlusion of other image regions did not have a comparable detrimental effect on performance. It had only minor effects on performance how much of an image was visible. Even when only 10% of an image were exposed, diagnostic stimuli were recognized. Also, in most cases non-diagnostic stimuli were not recognized despite the fact that half of the image was visible.

3.1.3 Partial occlusion of natural scenes - comparison between monkey and human observers

For a number of reasons it seemed interesting to determine the strategies of human observers in the paradigm used for the monkeys. A larger number of human observers could easily be tested, allowing to assess the variability of strategies between observers. This is especially interesting because the two monkeys were found to use very different strategies. In addition, testing more observers made it possible to determine the influence of some of the experimental parameters. Finally, a comparison between the strategies of humans and monkeys indicates whether these different observers are sensitive to the same kind of information in natural scenes when performing a task.

In the first experiment (Experiment 1), three human observers were trained to perform the Bubbles task on the first set of natural scenes. Since this experiment had actually been performed before testing the monkeys with natural scenes, it differed in a number of details from the experiment testing monkeys on the same images. For human observers, the stimulus set included four instead of three stimuli by adding a scene showing a pure landscape (scene N4). Also, stimuli had a smaller size of 128 by 128 pixel (4.2 deg). For the masks, a smaller bubbles diameter was chosen as well. A diameter of 0.95 deg was used (23% of the image size), so that despite

the differences in absolute size, the ratio between image size and bubble diameter was equal for monkey and human observers.

Human observers did not respond to the images by making a saccade; instead, they pressed one of four buttons on a computer keyboard. Stable eye movement recordings are more difficult to obtain from human observers. Since the monkeys used saccades simply to indicate the presented image, and there was no further interest in the properties of these eye movements, a change in response mode seemed uncritical. Each of the four buttons was associated with one of the images. Observers were first trained to respond correctly to the full scenes before they were introduced to the Bubbles paradigm. As for the monkey observers, images now appeared behind trial-unique masks with randomly positioned bubbles. As a control, unmasked trials were inserted into the testing, so that the performance on the basic task could be monitored. None of the subjects fell below 96% correct responses with the unmasked presentations.

About 1500 masked trials were collected per image in multiple sessions, and analyzed in the same way as the data recorded for the monkeys. Specifically, diagnostic regions were derived for this and the following experiments by subjecting the data from correct and incorrect trials to a Kolmogorov-Smirnov test, and thresholding the resulting p-values at .01, with a Bonferoni correction applied for the number of performed tests. The same threshold was applied to derive the diagnostic regions for monkeys G00 and B98 depicted in Figure 3.5 and 3.6. Figure 3.10A shows the diagnostic regions computed for the three subjects for each of the scenes.

Human observers were found to be robustly influenced by the occlusion of some regions in the natural scenes. This was the case for all of the scenes and all of the observers. Diagnostic regions always included parts of the object shown in a scene. In case of the landscape scene, which contained no such foreground object, observers identified the scene by the visibility of either the river in the middle, or the horizon. The fact that human observers consistently used object information to identify a scene distinguished their behavior from the one of the monkeys.

Since the same statistical threshold was used to compute the diagnostic regions for all observers, the size of the regions could be compared across observers. The extent of the diagnostic regions varied a lot between the human observers. While for example observer VB used information from rather small patches in the scenes, diagnostic regions for observer KN included at least half of each scene. Expressing the size of the diagnostic region as percent of the size of the entire image, the diagnostic regions of observer VB covered on average about 5% of the full stimulus size. For observer KN, diagnostic regions had an average size of 72% of the full stimulus.

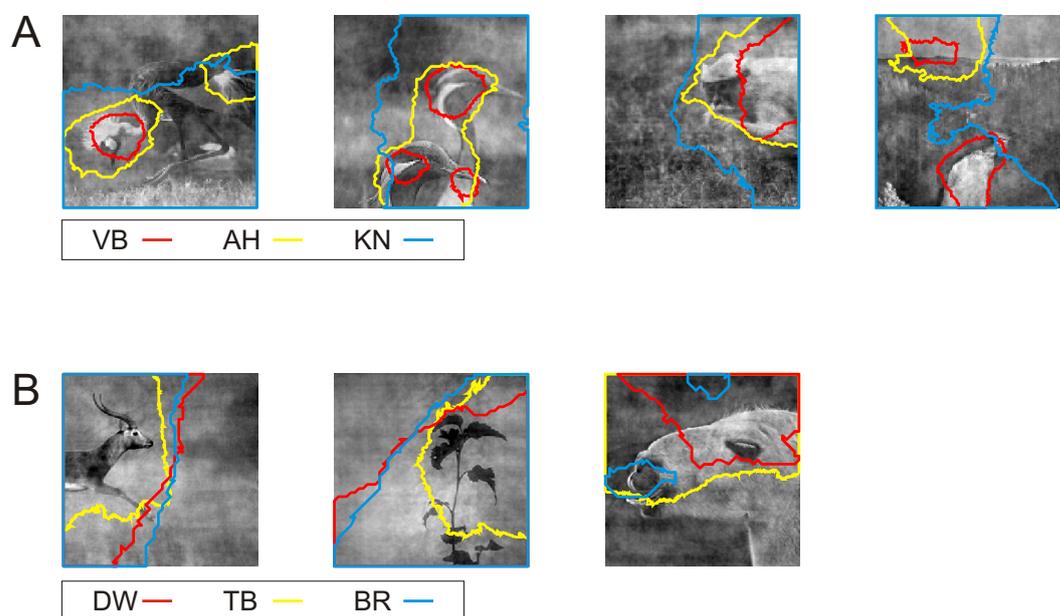


Figure 3.10: *Diagnostic image regions for human observers. A, Natural scene set 1 at size 128 by 128 pixel. B, Natural scene set 2. Lines enclose image regions with Bonferroni corrected p -values below .01. The results for each observer are plotted with a different color (see legends for the observers' initials).*

Despite this variability, human observers nonetheless used in general information from larger parts of the images than the monkeys. To compare the two observer groups, only images N1 - N3 were considered, since they had been shown to both groups. Averaging over images and observers, the diagnostic regions had an average size of 33.2% of the full image for human observers, and of only 5.0% for the monkeys. A t-test showed this difference to be significant ($t(13)=-2.75$, $p=.02$).

Another difference between monkey and human observers was found in the variability of decision strategies between observers: While the results of the two monkeys were very inconsistent, the diagnostic regions of multiple human observers mostly overlapped. The overlap was quantified as the size of the diagnostic regions shared between all observers, in relation to each observer's diagnostic region size. For the monkeys, diagnostic regions did not overlap at all. In contrast, the diagnostic region common to all observers explained about 52.0% of an individual subject's diagnostic region.

Three additional subjects were tested on the second image set (Experiment 2; scenes N5 - N7). In this case, the identical stimulus set was used for humans and monkeys; bubbles also had the same size on the screen. Since humans were placed closer to the screen than the monkeys, stimuli subtended a size of 8.3 deg for the human observers, and bubbles had a diameter of 1.9 deg. The results (see Figure 3.10B) confirmed the observations made for the first image set: The diagnostic regions of human observers always at least partially included foreground objects. Diagnostic regions were mostly large, with an average size across images and observers of 46.2% of the full image. They overlapped between different observers to a large degree, since on average 50.9% of a diagnostic region were common between all observers. In contrast, the monkeys' diagnostic regions had an average size of 8.6% of the full size, and overlapped only very little. On average, about 20.9% of an diagnostic region was shared between the monkeys; this overlap was mainly due to the overlap in diagnostic regions for image N5. The difference between the sizes of the diagnostic regions between monkeys and humans was significant ($t(13)=-4.25$, $p=.001$). The results of both experiments therefore showed the strategies of human observers to be more homogeneous, and to involve information from larger scene regions, than the strategies employed by the tested monkeys.

Besides the consistency and extent, the locations of the diagnostic regions also differed between monkeys and humans. Both monkeys used very different strategies for the two image sets. It seemed that the diagnostic regions of monkey G00 were more similar to the ones of humans, while monkey B98 had a very unique strategy. To quantify this effect, it was computed how much of a diagnostic region of a monkey could be predicted based on the human experiments. For this analysis,

the diagnostic regions of each human observer were compared against the diagnostic regions of one of the monkeys. Percentages were averaged across all observers and images. For monkey G00, human diagnostic regions predicted on average 82.7% of a diagnostic region for the first set of natural scenes, and about 81.2% for the second set of natural scenes. This shows that the smaller diagnostic regions for monkey G00 were mostly included in the diagnostic regions of human observers. In contrast, only 50.6% of the diagnostic regions of monkey B98 could be predicted from the human data for the first set of natural scenes. In the case of scene N2 alone, this value even dropped to 9%. For the second set of natural scenes, the human diagnostic regions predicted on average 60.4% of the monkeys behavior; for the three scenes, the values ranged between 17.8 and 99.0%. The locations of the diagnostic regions for monkey B98 thus only partially overlapped with the diagnostic regions of humans.

Experiments with human observers were also used to study the influence of two of the experimental parameters, namely image size and bubble size. These parameters were more closely inspected because both have the potential to influence an observer's strategy. Bubble size determines how large the image pieces are that are visible through each bubble. If a bubble does not permit enough image structure to be seen to resolve object parts, subjects may opt for using luminance and contrast differences to form their decision. On the contrary, with larger bubbles, objects may more easily be identifiable, and object information may therefore dominate in the scene identification task. The same arguments also hold for image size, since it determines the resolution of the presented image. Bubble size and image size could therefore bias an observer's strategy towards a more low-level or more high-level strategy (see Gosselin & Schyns, 2005, for a similar concern). Both questions were addressed in a third experiment involving six human observers. Images N1 - N4 were presented to the subjects, with a size of 256 by 256 pixels (8.3 deg). Three observers performed the Bubbles task with bubbles of 1.9 deg diameter (Experiment 3a), the other three with bubbles of 0.95 deg diameter (Experiment 3b). Image N4 consistently failed to evoke a sufficient number of error trials for the latter group of subjects; the statistical analysis was therefore not possible, and the image had to be excluded from the rest of the analysis.

Together with Experiment 1, the data from three experiments on the same image set could thus be compared. In Experiment 1, small images and small bubbles were used. Experiment 3a used large images and large bubbles, and Experiment 3b used large images with small bubbles. Figures 3.11A and B show the results of Experiment 3a and b. In addition, Figure 3.11C shows for each of the three experiments the common diagnostic region, i.e. the image region considered diagnostic by all subjects. For

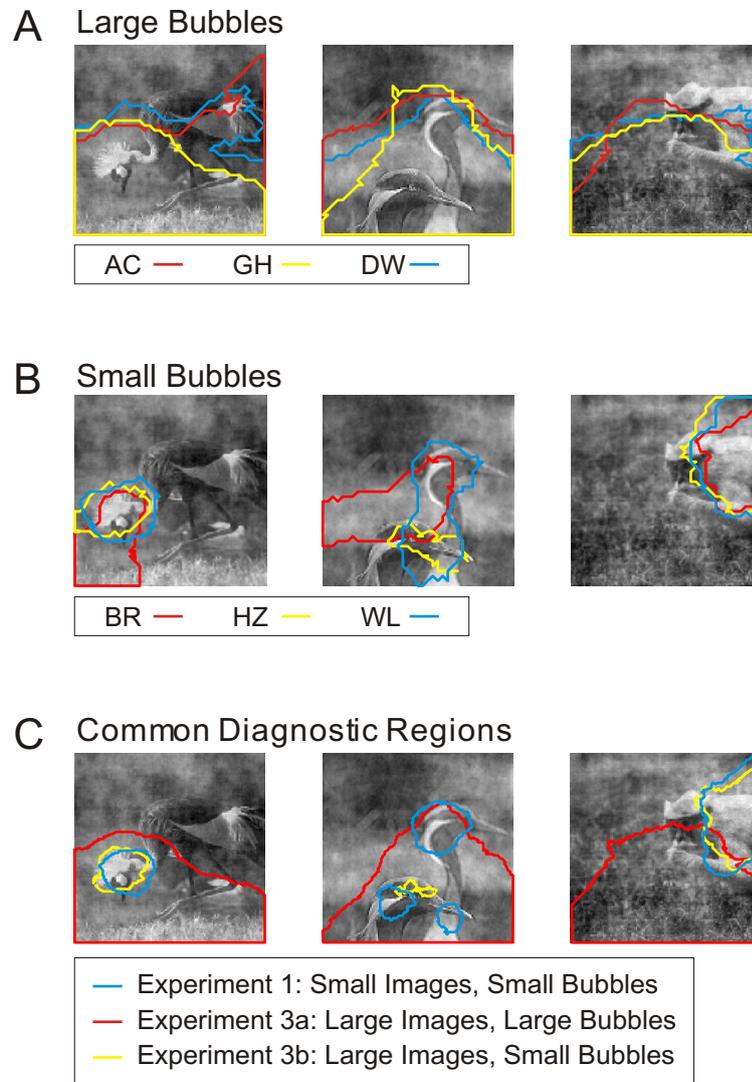


Figure 3.11: *Influence of bubble and image size. In contrast to the first experiment, images were shown with a size of 256 by 256 pixel. A, Diagnostic regions computed from an experiment involving masks with bubbles of 1.9 deg diameter. B, Diagnostic regions for masks with bubbles of 0.95 deg diameter. In A and B, lines again enclose regions of $p \leq .01$, and the color indicates the identity of the human observer. C, Common diagnostic regions, i.e. the diagnostic regions shared between all subjects, for Experiments 1, 3a, and 3b. The legend gives the line color for each experiment, as well as the experimental details. Small images had a size of 128 by 128 pixels, large images had the size of 256 by 256 pixels. For the bubbles, the small versions had a diameter of 0.95 deg, and the large ones of 1.9 deg.*

this plot, the data from Experiment 1 were rescaled to the image size of Experiment 3a and 3b.

Experiment 1 and 3a shared the same ratio between bubble diameter and image size. Stimuli in Experiment 3a were thus simply scaled versions of the stimuli in Experiment 1. Most importantly, the amount of image structure visible through a single bubble remained unchanged. Nonetheless, the data showed that increasing the stimuli resulted in larger diagnostic regions. For Experiment 3a, the diagnostic regions covered on average 56.9% of the scenes, which was significantly larger than the average diagnostic region size of 33.2% in Experiment 1 ($t(16)=2.83$, $p=.01$). One subject in Experiment 1 had very large diagnostic regions as well. This shows that in principle more information could have been used in the smaller stimuli. However, it required the larger stimuli for this information to be consistently used by all subjects. Considering the positions of the diagnostic regions, it seems that information especially in the background regions became more easily accessible in the larger stimuli. This hypothesis was confirmed by splitting each scene into background and foreground regions. The overlap between the diagnostic regions and the foreground regions was then computed. Indeed, foreground regions constituted 56.3% of the diagnostic regions in Experiment 1, and only 27.5% in Experiment 3a ($t(16)=-4.11$, $p=.001$). Thus, the background region conveyed more information in the larger stimuli than in the smaller ones.

Since both image and bubble size were doubled, these effects cannot be explained by a bubble size that only revealed object information at a certain scale in Experiment 1. Instead, the reason for the larger diagnostic regions has to be found in the fact that small detail became more visible in the larger stimuli. This is consistent with the fact that the image regions that were only diagnostic with larger stimuli corresponded to background regions containing lots of small detail.

Figure 3.11C shows that when increasing the image size without increasing the bubble size, diagnostic regions remained relatively unchanged. Diagnostic regions became moderately smaller in Experiment 3b than in Experiment 1 (average size of 14.1% of the image size, instead of 33.2% for Experiment 1; $t(16)=-2.29$, $p=.04$). However, the location of the diagnostic regions was similar for Experiments 1 and 3b. The common diagnostic regions for Experiment 1 explained on average 63% of the diagnostic regions of Experiment 3b. Excluding scene N2 because of the bad agreement between observers in Experiment 3b even increased this value to 89%. This also implied that diagnostic regions were still centered on object regions, and not on background regions with prominent low-level characteristics.

In Experiment 3b the ratio of bubble diameter and image size was changed relative to Experiment 1. Image features were thus visible at a smaller scale. Nonetheless, observers' strategies remained relatively unchanged. The number of bubbles is flexibly adjusted to an observer's performance in the Bubbles paradigm. If image material is presented at a scale that is too small for an observer, Bubbles can accommodate by increasing the amount of bubbles. A comparison of the bubble numbers between Experiments 3a and b shows that this mechanism might indeed be responsible for the overall similar diagnostic regions.

During all experiments involving human observers, bubble numbers were adjusted to a subject's performance using an automated staircase protocol so that the subject responded correctly about 75% of the time. Staircases quickly reached their asymptote. The last 100 trials of each session performed by a subject could thus safely be taken as a measure of the number of bubbles necessary for a performance at this level. For Experiment 3a and b, subjects tested with small bubbles required different bubble numbers than subjects tested with large bubbles. With large bubbles, staircases resulted in an average of 3.0 bubbles (averaged over all sessions from all three observers). In contrast, 11.3 bubbles were necessary for subjects tested with small bubbles to reach the same performance level. This difference was highly significant ($t(32)=8.94, p<.000$). The data show that observers working with masks punctured by small bubbles required about four times as many bubbles as subjects tested with large bubbles. Since the bubble diameter was doubled from small to large bubbles, this implied that the average area covered by the bubbles – i.e. the stimulus area visible through the masks – was almost identical for the two groups. For an equal performance, subjects from both groups therefore needed to be exposed to about the same proportion of the occluded stimulus.

In conclusion, diagnostic regions were not affected by changing the scale at which image information was available. Combining the results of Experiments 1, 3a and 3b, changes in the relation between bubble size and image size seemed not to affect the observer's strategy; possibly because the bubble number was increased for the smaller bubbles. Importantly, there was no indication of a more low-level strategy with smaller bubble sizes. Increasing the image size allowed subjects to use additional information in the images. This effect is probably due to the higher resolution of the larger images.

One more experiment was performed with human observers which served to verify the Bubbles results. Some of the subjects who had performed in the experiment before were provided with printouts of the experimental stimuli after the final testing session. They were asked to encircle stimulus regions that they had found diagnostic during the task, and to list any additional strategy that they had used. Figure 3.12

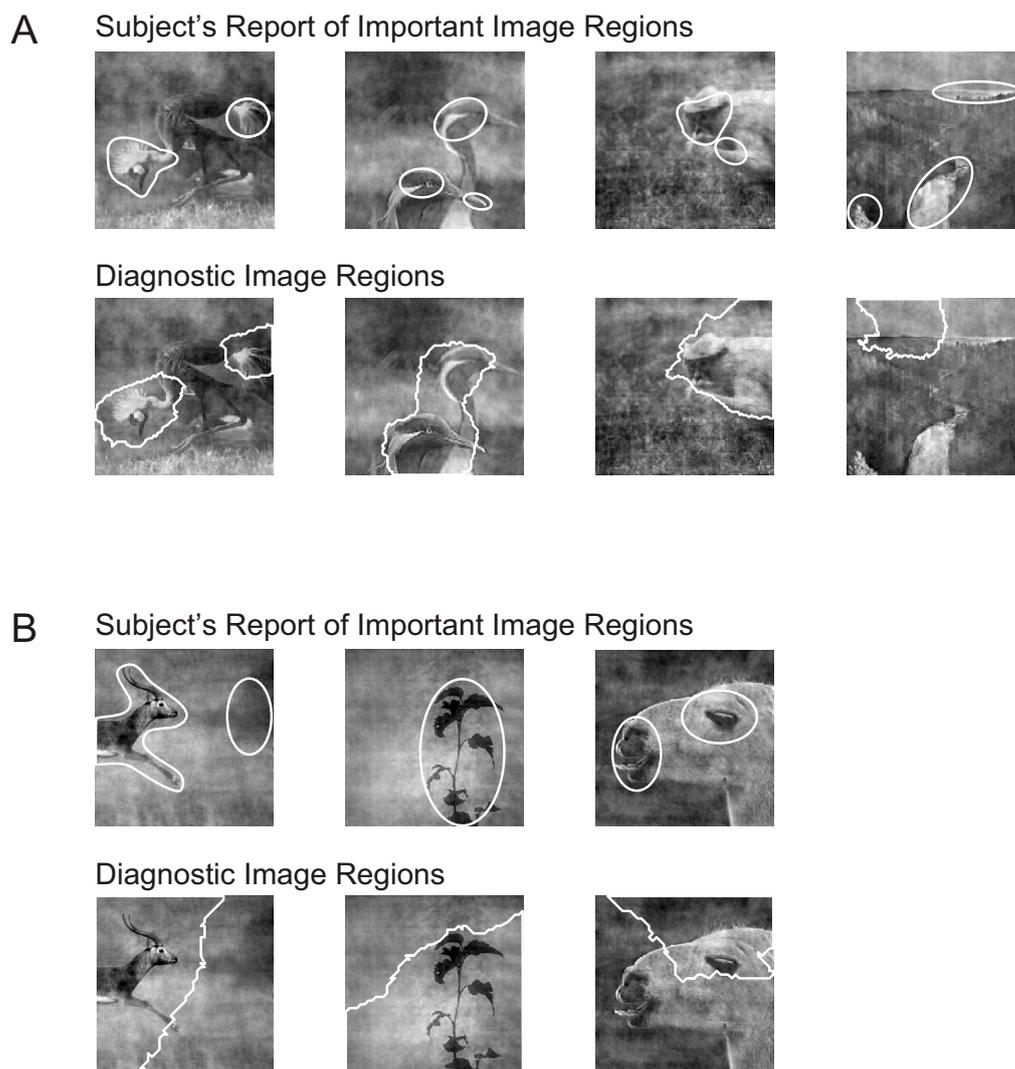


Figure 3.12: Comparison between the diagnostic regions determined with *Bubbles*, and an observer's insight into his strategy. *A*, Results from observer *AH*. *B*, Results from observer *DW*. The top row in each panel shows the regions indicated as diagnostic by the subject, the bottom row gives the *Bubbles* results.

shows a comparison of the subjective reports and the Bubbles results for two subjects. In most cases, the subjective reports were in good agreement with what had been determined as diagnostic using Bubbles. Differences most likely were generated by how consistently subjects used information from the image regions. In cases in which they overestimated their consistency, regions appeared in the subjective reports but were absent in the Bubbles results; most likely because they failed to reach significance. In other cases, subjects underestimated how much information they actually gained from a region. In these cases, diagnostic regions were larger for the Bubbles results than for the subjective reports. To summarize, the subjective reports confirmed that the Bubbles results indeed reflected an observer's strategy to solve a task.

3.1.4 Eye movements versus Bubbles

Visual acuity is highest in the central 2 deg of vision, around the fovea. Human and monkey observers make use of this fact, and reorient their center of gaze around a viewed scene at a rate of about 3 times per second. A very stereotyped eye movement behavior is evoked in these situations: The gaze is moved from one image position to another by a very quick eye movement, the so called saccade. After a new position has been acquired, the gaze remains almost still at this position, until it is reoriented again by a saccade. The scanning behavior therefore consists of a sequence of alternating fixation periods and saccades. Pattern information is acquired during the fixations, while visual sensitivity is reduced during a saccade (Ross et al., 2001).

One of the earliest studies examining where human observers fixate when inspecting a complex scene was performed by Yarbus (1967). He reported that when viewing a painting of a realistic scene, human observers mostly fixated on people present in the scene, and particularly on the peoples' faces. The way in which viewers scanned the scene could be influenced by giving the subjects instructional questions before scene investigation (see Figure 3.13). Yarbus therefore suggested that the eyes tend to land on regions containing information that is "useful or essential for perception". This hypothesis was tested more formally by Mackworth & Morandi (1967) and Antes (1974), who studied whether fixations preferably fall onto informative scene parts. In both experiments, scenes were divided into smaller patches. One group of subjects was asked to rate the informativeness of each of these patches with respect to the information contained in the large scene. Subsequently, the eye movements of a second group of observers were monitored while they inspected the

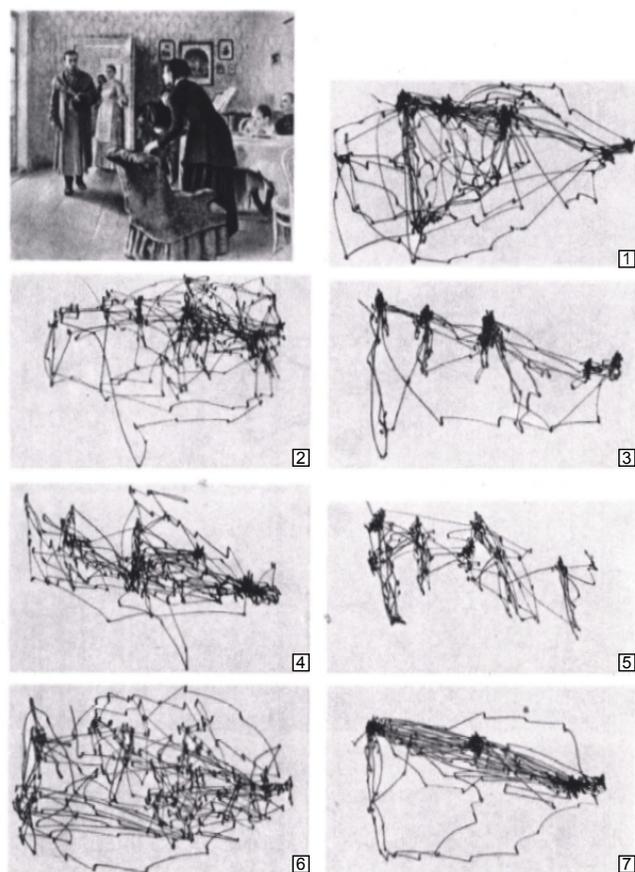


Figure 3.13: *Examples of eye scanning records obtained by Yarbus (1967). Observers were given different instructions while viewing the picture “They did not expect him” by Ilya Repin. Each of the traces shows a 3 min record of eye scanning with the following instructions: 1) Free examination. 2) Estimate the material circumstances of the family. 3) Give ages of the family. 4) Surmise what the family had been doing before the visitor’s arrival. 5) Remember the clothes worn by the people in the picture. 6) Remember the position of the people and objects in the room. 7) Estimate how long the “unexpected visitor” had been away.*

large scenes. Regions rated as highly informative attracted the most fixations, while region considered to be uninformative were often not fixated at all.

The two studies described above based their conclusions on the differences in informativeness between image patches. However, the image patches did not only differ in how much information they contained, but also in many of their low-level properties. In principle, each of the two factors alone may have been sufficient to influence the distribution of fixations. The influence of pure semantic inconsistencies, i.e. the

influence of informativeness alone, was studied in a number of studies. Recently, extending a study originally performed by Loftus & Mackworth (1978), Henderson et al. (1999) manipulated the consistency of particular regions of complex, natural scenes by changing a target object in that region. Low-level information was controlled by pairing scenes, and exchanging objects across scenes. As an example, a laboratory and a bar scene were paired so that either scene could contain a microscope or a cocktail. Depending on the background scene, the same object was either consistent or inconsistent. A rectangular region was defined around each target object, and the number of fixations within or outside of the region was measured. In agreement with the reports by Mackworth & Morandi (1967) and Antes (1974), the results showed observers to fixate more often on inconsistent than on consistent targets. This effect was due to a longer looking time on the inconsistent region, as well as more returns of the gaze to this region. No influence of semantic consistency was evident in the initial fixations; inconsistent targets, for example, were not fixated earlier than consistent ones.

It is still an open question whether the image regions attracting a fixation can be distinguished based on some of their low-level, physical properties (Reinagel & Zador, 1999; Mannan et al., 1996). It has, however, been demonstrated that these properties control at least part of the scan patterns of human observers. Mannan et al. (1995) presented observers either with unfiltered natural scenes, or with the same scenes strongly low-pass filtered. The low-pass filtered scenes had the same overall spatial layout as the original scenes, but were not identifiable by the observers. Nonetheless, the distribution of fixations across the scenes was similar for both scene types, especially during the first 1.5 s.

The currently prevalent view about the factors determining the placement of fixations on a natural scene can be summarized as follows (Henderson & Hollingworth, 1999): Upon presentation of a complex scene, the initial fixations are mainly driven by the low-level characteristics of a scene in a bottom-up manner. During this early phase, fixation locations are independent of the content of a scene; later fixations, however, are influenced by the nature of a scene and task requirements. As a consequence, the spatial distribution of fixation density is controlled by the how information is distributed across a scene, and informative regions attract in total more fixations than non-informative regions.

The fixation density observed when a viewer scans a scene can therefore potentially reveal which scene regions the observer finds interesting. So far, Bubbles has been used to characterize how information is distributed across natural scenes. Now, the fixation density on the same scenes was recorded for the monkey observers under two different conditions. In the first experiment, scenes were simply presented to the

monkeys for free-viewing. No particular response was required from the monkeys, so that this completely unconstrained situation serves as a baseline to highlight image regions that the monkeys in general find interesting. The evoked fixation patterns also give an indication of whether the monkeys were able to parse the stimuli into foreground and background, resulting in meaningful images. The Bubbles results, however, were obtained while the monkeys had to actively discriminate between different images. Especially the results from monkey B98 suggest that under task conditions, very specific image regions attain diagnosticity. The same regions may be completely uninteresting outside the task. It was therefore tested whether the fixation density observed while the monkeys performed a task on an image changed with respect to the free-viewing condition, and how it aligned with the Bubbles results.

All eye movements of the monkeys were recorded using a scleral search coil, which allowed the position of the gaze to be detected very accurately. Before any scanning data was obtained, the eye position recordings had to be calibrated in separate sessions. During calibration, the monkeys fixated a small square presented on various locations on the screen. Each location was tested repeatedly. For each fixation, the difference between the required and the actual fixation location was determined. The absolute value of this difference was averaged over repetitive tests at the same location. The average gave a measure for the accuracy with which fixations could be attributed to a specific spatial location. It reflected both the measurement precision of the setup, as well as the accuracy with which the monkey was fixating. Initially, accuracy was determined as a function of the spatial location of a fixation on the screen. It was, however, found to be independent of the spatial location for both monkeys, and the data from different locations were pooled. General values for the position error in horizontal and vertical direction were thus computed. Horizontal values yielded 0.22 deg for monkey G00, and 0.26 deg for monkey B98. In the vertical direction, fixations could be attributed to a specific spatial location with an error of 0.26 deg for monkey G00, and 0.35 deg for monkey B98.

After the eye movement signal had been calibrated, the monkeys were tested either on a free-viewing or on a discrimination task. Both monkeys participated in the free-viewing task, but only monkey B98 was tested in the discrimination task. Data were collected before as well as after the Bubbles testing. During free-viewing, the scenes from one of the stimulus sets were shown to the monkey in pseudo-random order, who was free to inspect them. Each stimulus was shown for 3 s, so that there was ample time for eye movements. No constraints were placed on the monkey's behavior, and no reward was given, so that it was impossible to form an association between a particular behavior and a reward. The experiment was continued as long

as the monkeys were willing to look at the stimuli. Usually, between ten to 15 trials, i.e. three to five repetitions of each stimulus, could be collected. Experiments were repeated on multiple days to collect enough data.

During each trial, the eye position was continuously monitored. Fixation periods were extracted from the continuous traces by identifying time periods in which the eye movement velocity did not exceed a set threshold. For each fixation, the location was determined as the average gaze position during the fixation duration. Since the calibration data showed fixations to be scattered around a location that the monkeys intended to fixate, possible variability was taken into account by attributing a fixation not only to a single spatial location. Instead, locations were assigned to a rectangular region centered on the fixation location, with a width and height determined as twice the horizontal and vertical error, respectively. Fixation density could then be computed for each pixel of the tested scenes by counting the number of fixations falling onto this pixel, combining the data from all relevant trials in all sessions. Fixation counts were finally converted into the probability of observing a fixation on an image pixel by normalizing the counts to the total number of fixations falling onto an image.

By these means, the fixation density could be computed for each of the scenes (see Figure 3.14). The results showed that when freely inspecting a visual scene, both monkeys behaved very similar: When animals were present in a scene, the monkeys preferentially inspected the animals' heads, as well as their eyes. The only distinction seemed to be stimulus N1, which evoked very distributed fixation patterns. For scene N7, which contained no animal, the monkeys mainly fixated on parts of the plant visible in the foreground. These scan patterns are very consistent with the behavior of human observers inspecting a real-world scene. When inspecting real-world scenes, humans mostly fixate on the objects present in the scene; uniform background regions receive little or no fixations (Yarbus, 1967; Henderson, 2003). Furthermore, a number of studies reported that monkeys preferentially fixate on the major face features like eyes and nose when viewing facial displays. From these face features, the eyes receive most of the fixations (Wilson & Goldman-Rakic, 1994; Guo et al., 2003; Nahm et al., 1997). Again, our data are in agreement with these previous studies. Whenever a face was present in a scene, most fixations landed on the eyes. The results therefore did not only establish which scene regions the monkey attended to in general. Since the monkeys showed the typical behavior when inspecting scenes containing animals, the data also implied that the monkeys were able to perceive the content of the scenes correctly.

The influence of a task on the allocation of fixations was subsequently assessed by measuring eye movements during a discrimination task. As mentioned before, only

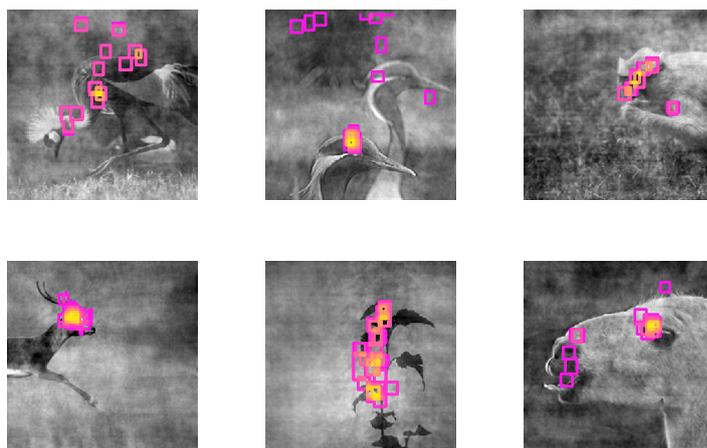
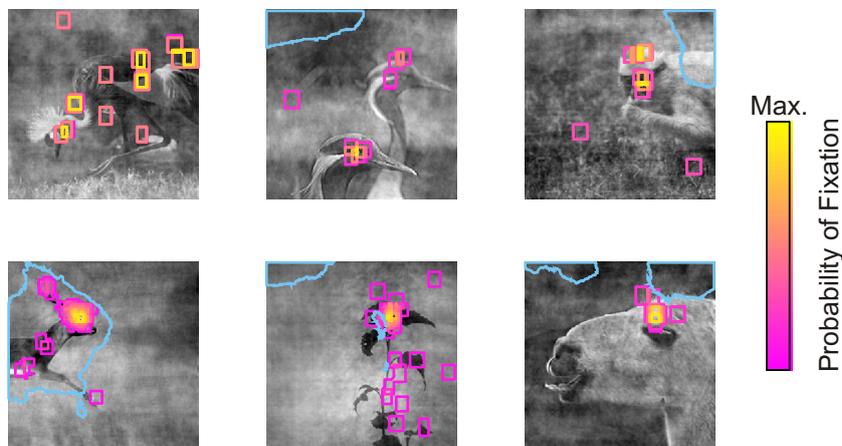
A G00, Fixation Locations During Free-Viewing**B** B98, Fixation Locations During Free-Viewing

Figure 3.14: Fixation density in a free-viewing paradigm. Contour plots indicate the probability of a fixation falling into an image region. Each contour plot is individually scaled so that the maximal value is plotted in yellow. A, Results from monkey G00. B, Results from monkey B98. The light blue lines in B indicate the positions of the diagnostic regions determined with Bubbles.

monkey B98 was tested in this experiment. The task was kept as similar as possible to the usual discrimination task that the monkeys were trained to perform. As usual, each trial began when the monkey fixated a centrally presented fixation spot. The spot was replaced by one of the stimuli, and each trial ended with the presentation of three peripheral targets, one of which the monkey had to saccade to. While in the normal trials the monkey was required to maintain fixation in the center of the screen during scene presentation, scanning trials were now introduced in which the monkey was free to move its gaze during the scene presentation. All other fixation requirements were kept the same. Scanning trials were mixed with normal trials and indicated to the monkey by placing a large rectangle around the fixation spot that initiated each trial. Correct responses in both trial types were rewarded by a drop of juice. In normal trials, stimuli were presented for 300 ms only. Since this is on the order of the duration of a single fixation, these trials did not provide enough time to evoke a sufficient number of eye movements. In scanning trials, the stimulus duration was therefore increased to 3 s, and only these trials were considered further.

Again, eye movements were recorded continuously during each trial. As before, fixation periods were extracted using a velocity criterion, and the probability of a fixation was computed for each image region. The monkey was required to fixate the fixation spot before the onset of a scene, and needed some time until the first saccade was executed. The first fixation after stimulus onset was therefore usually located at the center of the screen, and was not part of the scan pattern evoked by an image. It was therefore excluded from the analysis. The resulting distribution of fixations can be seen in Figure 3.15. A major difference between this data and the data collected during free-viewing has to be taken into account when interpreting any effects: In the free-viewing condition, the monkeys were not performing a task, and they were not rewarded. As a consequence, they usually quickly lost interest in the scenes, and stopped to examine them further. In contrast, in the the second experiment the performance of a task required more constant attention. Even though multiple sessions were combined for the free-viewing data, the data set from the task condition was still much richer and contained many more fixations.

Nonetheless, a number of observations were made by comparing the fixation density from free-viewing and task conditions. During the task, the main fixation locations were the same as during free-viewing. The monkey continued to fixate on the head of the displayed animals, and in particular on their eyes. Maxima of the fixation densities had an average distance of 25 pixels between task and free-viewing maps, slightly less than 10% of the image size. Both data were therefore in good agreement with respect to the regions receiving most of the fixations.

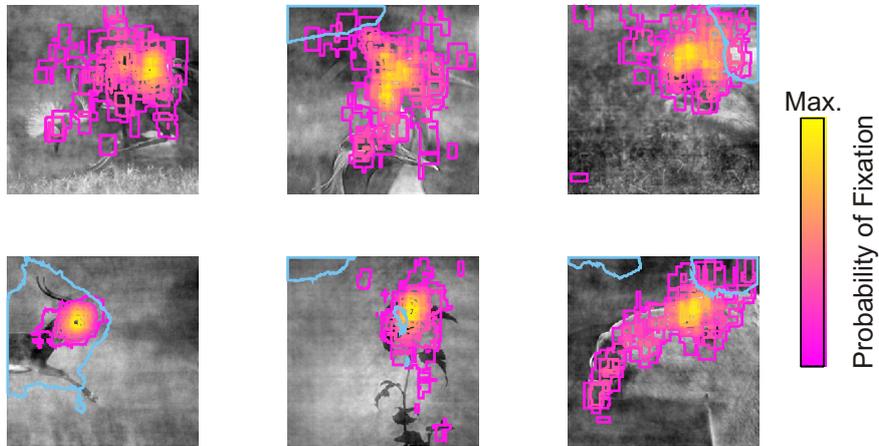


Figure 3.15: *Fixation density during a discrimination task. As before, contour plots indicate the probability of a fixation falling into an image region. Results for different images are scaled individually. Again, the blue lines indicate the diagnostic regions determined with Bubbles.*

Possible differences between fixations in free-viewing and task conditions emerged when comparing the fixation densities to the Bubbles results. This was particularly the case for the first set of natural scenes. Here, the Bubbles results revealed diagnosticity mainly to be concentrated in background regions (see Figure 3.6A). These regions did not receive fixations in the free-viewing task, but some proportion of fixations during the discrimination task. This proportion was quantified as the percentage of fixations in the Bubbles diagnostic region. Only scenes N2 and N3 were considered because of the lack of an appropriate diagnostic region in N1. While in the free-viewing task 0% of the fixations fell into the diagnostic region, during the discrimination task 10% of the fixations were placed into the diagnostic region. For the second set of natural scenes, the diagnostic regions covered more object parts (see Figure 3.6B). For this stimulus set, the discrepancy between fixation locations in the free-viewing task, and diagnostic regions was therefore not as large. Consequently, the diagnostic regions received about as much fixations in the free-viewing and the discrimination task (on average, 38% in the free-viewing, and 34.5% in the discrimination task). The comparison between free-viewing and task data may be confounded by the different amounts of fixations in the two experiments. However, at least for the first set of natural scenes there was a trend for more fixations in the diagnostic regions during the performance of a task. It is possible that this trend is

generated because the monkeys use information from this region during the task, as indicated by the Bubbles results.

In conclusion, the fixation patterns obtained when monkey observers viewed the natural scenes confirmed that the monkeys were able to extract the meaning of the scenes, in the sense that they seemed to react appropriately to the depicted animals. Fixation patterns obtained during a task were mainly consistent with the fixation patterns recorded during free-viewing; however, there may have been an influence of the task on the distribution of fixations across different image regions. Nonetheless, even when recorded during a task, regions with maximal fixation density could not be used to predict the diagnostic regions found with Bubbles.

3.1.5 Characterization of diagnostic image regions

Diagnostic image regions have so far been described in terms of their location, as well as their extent. In addition to these parameters, the physical properties of these image regions were also analyzed. Consistent differences between diagnostic and other image regions may help to pinpoint why observers relied on these regions to perform the task. For this purpose, the luminance and edge energy of individual image regions were first computed. These two parameters were chosen because they represent image characteristics that neurons in early visual areas are known to be sensitive to (Hubel & Wiesel, 1977). Both parameters were extracted at four spatial resolutions by applying filters of a fixed size to scaled versions of an image. For the highest resolution, images had their original size of 256 by 256 pixel. With each scaling step, the image size was reduced by a factor of four by bisecting the side length of the image. Luminance information was computed by convoluting the scaled images with a 2-D Gaussian, while four pairs of oriented Gabor filters, each pair with a different orientation, were used to calculate edge density (see Figure 2.16 for examples of filtered images). Each feature map was rescaled to half the image size, irrespective of the image size used during the filtering process. For both parameters, the maps at different spatial resolutions were combined following a model by Itti et al. (1998) to form maps representing more higher-order information. These maps depicted the saliency of luminance or edge information at a particular image location. The saliency maps computed for both sets of natural scenes are shown in Figure 2.17. In total, four luminance, 16 edge, and two saliency maps showing the saliency of luminance and edge features were generated for each image. Since the absolute values of these maps were not of interest, map values were rescaled to span the range from 0 to 1.

Using these maps, the properties of the diagnostic regions could be computed and compared to other image regions. In most cases so far, diagnostic regions were defined as the image region with a p-value of less than .01 in the statistical maps. This definition was not practical here for two reasons: First, the properties of diagnostic regions needed to be contrasted with other image parts, which had to be of the same size, and should be non-overlapping with the diagnostic regions. The usual criterion did not easily lead to an objective rule for the construction of control regions that fulfilled these criteria. In addition, using a criterion like $p \leq .01$ generated diagnostic regions with varying size, introducing unwanted variability in the results. A different method was therefore chosen to define diagnostic regions. Here, image pixels were included in the diagnostic regions if their p-values belonged to the 30% smallest p-values. Non-overlapping control regions of the same size could then simply be generated by selecting image regions with p-values belonging to the 30% largest p-values. A similar approach has already been used to construct controlled masks to verify the Bubbles results for the monkeys (see Section 3.1.2). A size of 30% was chosen so that the resulting diagnostic regions were not too small. Changing the size of the diagnostic regions to 10% of the image size did, however, not affect the results.

The properties of the diagnostic and non-diagnostic regions were then quantified by averaging the normalized feature values from all pixels within these regions. Spatial resolutions were treated separately, but the four orientation maps at each resolution were averaged. Figures 3.16 and 3.17 show the average values for luminance, edge energy, luminance saliency, and edge saliency in diagnostic and non-diagnostic regions for monkey and human observers. To generate the plot for the monkeys, the results from both experiments involving natural scenes were combined, i.e. each monkey contributed six diagnostic regions and their control regions. For the human results, the data collected from all subjects tested with stimuli of a size of 256 by 256 pixels were combined. These results were therefore based on 27 diagnostic regions from nine observers.

For the monkeys, no consistent differences were found between diagnostic and non-diagnostic regions in terms of their average luminance at any of the scales (paired t-tests at each of the resolutions; $t(11) \leq 1.23$, $p > .24$ at all spatial resolutions). The same was the case for both luminance and edge saliency. Diagnostic regions could not be distinguished from non-diagnostic regions based on these parameters ($t(11) = 1.31$, $p = .20$ for luminance saliency, $t(11) = 0.19$, $p = .85$ for edge saliency). However, diagnostic regions contained more oriented edges than the non-diagnostic regions, at least when edge energy was computed at the coarsest resolution ($t(11) = 3.62$, $p = .004$). Differences were not significant for the three finer spatial resolutions

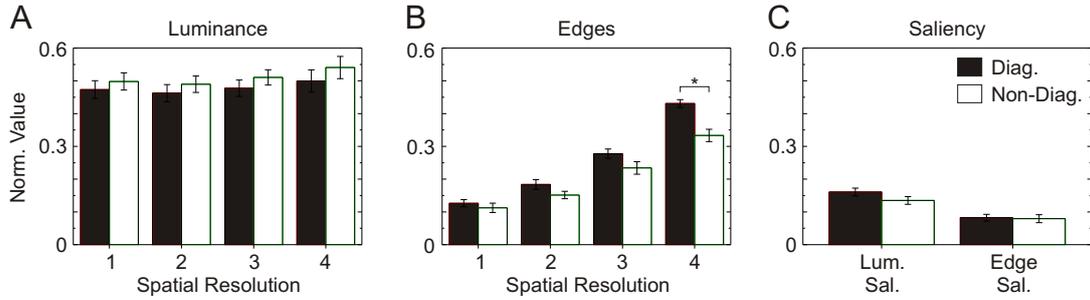


Figure 3.16: Physical properties of the diagnostic regions for monkey observers (black bars). The properties of non-diagnostic regions are shown as a control (white bars). Stars indicate differences between diagnostic and non-diagnostic regions significant at $p \leq 0.05$. A, Luminance distribution. B, Edge information. C, Saliency maps. Error bars correspond to the standard error of the mean.

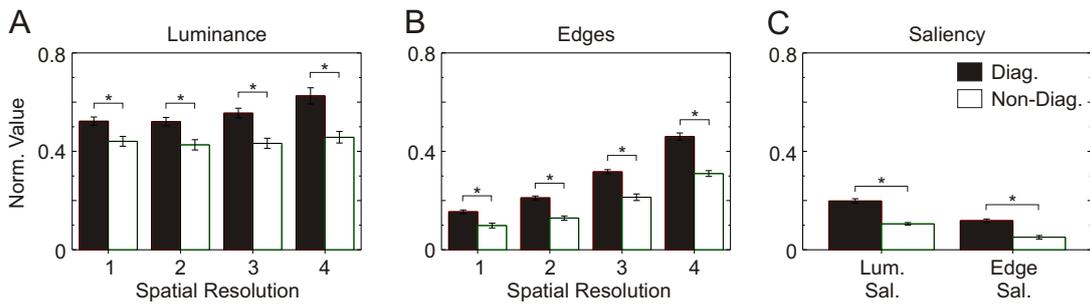


Figure 3.17: Physical properties of diagnostic and non-diagnostic regions for human observers. Notation is identical to Figure 3.16.

($t(11) \leq 1.76$, $p \geq .11$ for resolutions 1 to 3). In contrast, diagnostic image regions of human observers consistently differed from the non-diagnostic regions in all computed parameters. For all paired comparisons, these differences were highly significant ($t(26) \geq 3.41$, $p \leq .002$ for all comparisons).

Since some image features were found to consistently differ between diagnostic and control regions, it was tested whether the pronounced presence of some of the image features could be used to predict the diagnostic image regions. In each of the feature maps, the locations were identified where the feature values belonged to the highest 30% of the values (the “feature regions”). Feature regions therefore had the same size as the diagnostic regions used here, and their locations could directly be compared to the locations of the diagnostic regions. If, for example, an observer identified a scene by its bright regions only, the luminance feature region would correspond to the diagnostic region. Feature regions were defined as regions with high values for luminance, edge, and both saliency maps. Since dark image regions might be as diagnostic as bright image regions, the image regions with luminance values belonging to the lowest 30% were additionally identified. The overlap between these feature regions and the diagnostic regions was then computed, to measure which percentage of the diagnostic region could be attributed to the pronounced presence of one of the features. As a comparison, the same was done for the non-diagnostic regions.

The performance of the physically defined image properties was additionally contrasted with how well the presence of an object, or an animal’s head, could predict the diagnostic regions. For this analysis, the image part covered by the foreground object was determined. These object regions represented on average 23% of a scene, therefore being slightly smaller than the feature regions. For all scenes containing an animal, the location of its head was also identified. For scene N6, the upper part of the plant was outlined. Obviously, the head region was smaller than the other regions defined so far, covering on average 14% of a scene. The overlap between the object or head region and the diagnostic and non-diagnostic regions was subsequently computed.

Results can be found in Figure 3.18 and 3.19. It is evident from these plots that for none of the parameters, image regions with extreme values in this parameter covered more than 50% of the diagnostic regions. This was the case both for monkey and human observers. For the monkey observers, a comparison between diagnostic and non-diagnostic regions showed that regions with high luminance in the images had no predictive power for the diagnostic regions (paired t-tests, $t(11) \leq 0.2$, $p \geq .84$ for all resolutions). Dark image regions overlapped somewhat more with diagnostic than non-diagnostic regions, except for the coarsest resolution; however the differences at

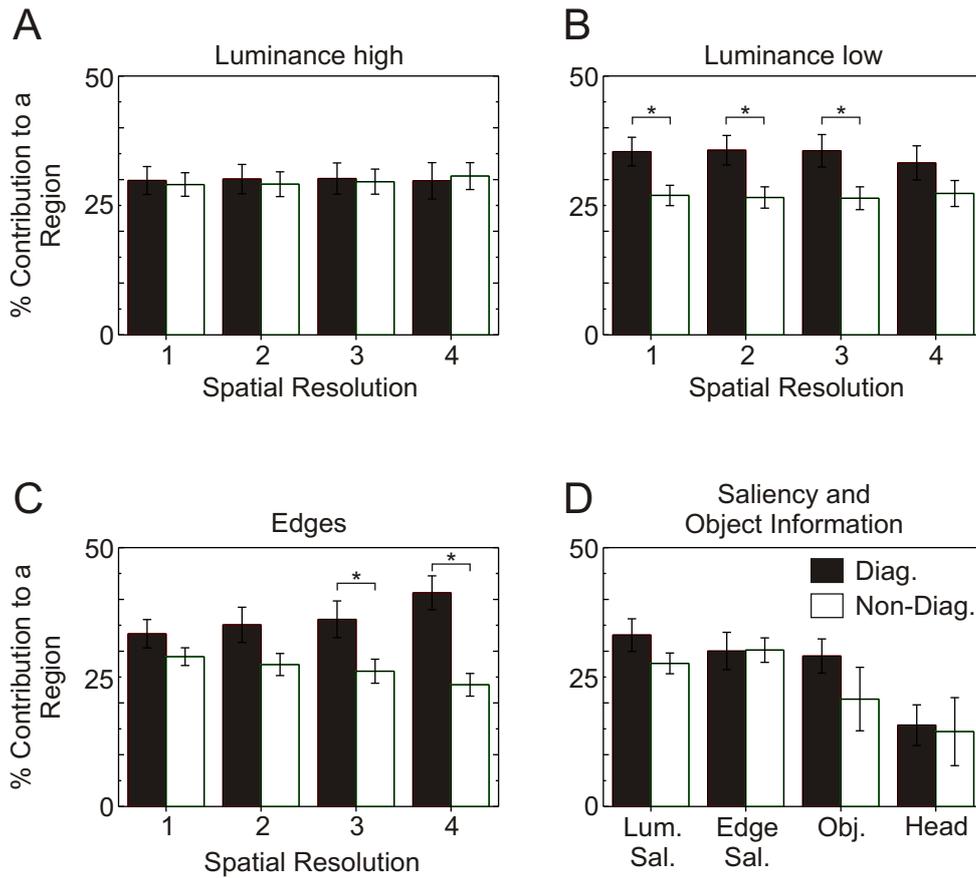


Figure 3.18: *Overlap between the image regions in which a feature is maximally expressed (feature regions), and the diagnostic regions of monkey observers (black bars). White bars indicate the overlap between feature regions and non-diagnostic regions. Stars indicate differences between diagnostic and non-diagnostic regions significant at a level of $p \leq .05$. A, Overlap with image regions with high luminance values. B, Overlap with dark image regions. C, Overlap with feature regions computed from edge density maps. D, Overlap with salient or object regions. Error bars show the standard error of the mean.*

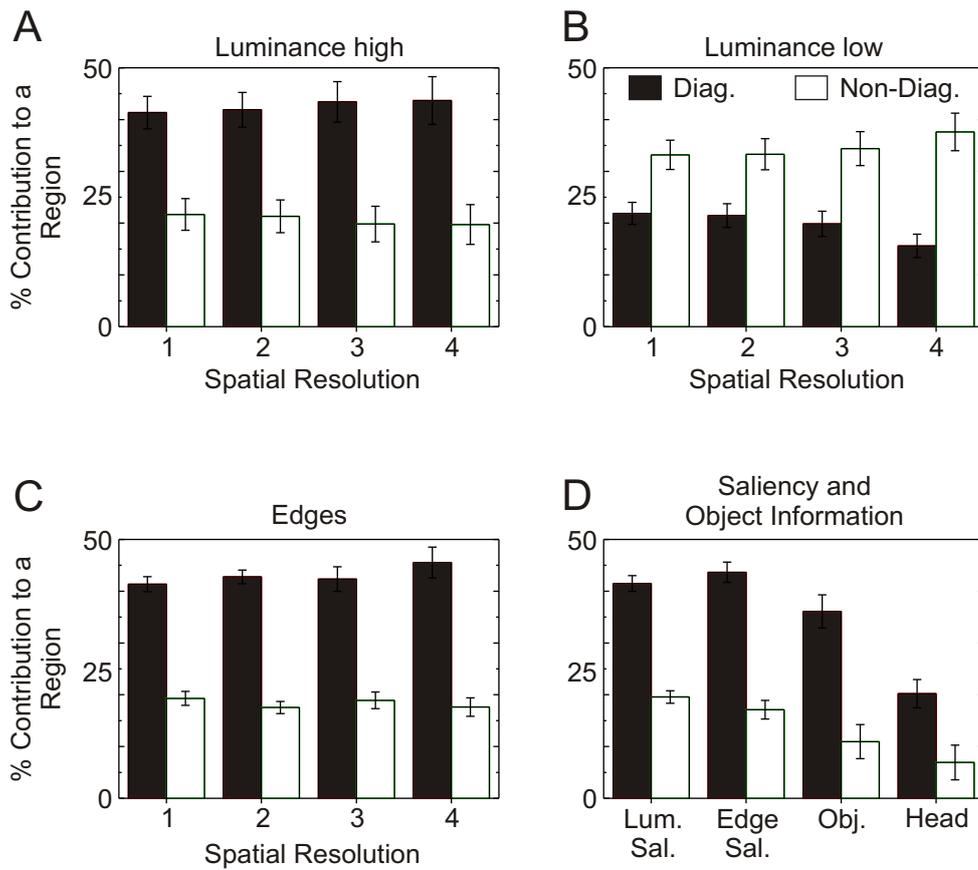


Figure 3.19: Portion of the diagnostic (black) and non-diagnostic (white) regions of human observers that could be explained by the feature regions. All differences between diagnostic and non-diagnostic conditions are significant ($p \leq .05$). See Figure 3.18 for further details.

the finer resolution also only just approached significance (resolution 1: $t(11)=1.95$, $p=.08$; resolution 2: $t(11)=2.00$, $p=.07$; resolution 3: $t(11)=1.85$, $p=.90$; resolution 4: $t(11)=1.09$, $p=.3$). In contrast, edge information seemed to explain more of the diagnostic regions only when computed with a coarse resolution (resolution 1: $t(11)=1.04$, $p=.32$; resolution 2: $t(11)=1.45$, $p=.18$; resolution 3: $t(11)=1.77$, $p=.1$; resolution 4: $t(11)=3.32$, $p=.007$). Neither luminance nor edge saliency could predict diagnostic regions better than non-diagnostic regions ($t(11)=1.10$, $p=.29$ for luminance saliency; $t(11)=-0.03$, $p=.97$ for edge saliency). In contrast, for human observers almost any feature region could explain more of the diagnostic than the non-diagnostic regions ($t(26)\geq 3.01$, $p\leq .006$ for all comparisons of high luminance, edge, and saliency regions). Dark image regions consistently overlapped more with the non-diagnostic than the diagnostic regions.

The overlap between object and diagnostic regions was as large as the overlap between the feature and the diagnostic regions. This was the case although the object regions were on average smaller than the defined feature regions. The head region explained less of the diagnostic regions than other features; however, these regions were much smaller than any other region.

In summary, the results showed that for the monkeys, diagnostic regions differed from non-diagnostic regions only in their edge content. For humans, each of the parameters distinguished between diagnostic and non-diagnostic regions. Despite these differences, the diagnostic regions could not be completely explained by how certain physical image features were distributed in the natural scenes. In addition, no feature exceeded the others in terms of how much it could predict the diagnostic regions. Simply using the location of a foreground object worked equally well to predict the location of the diagnostic regions.

3.2 Physiology results

The first two parts of this section describe the effects of stimulus occlusion on the responses of single neurons in area TE, as well as on the LFP responses. Finally, the single unit behavior will be compared to the LFP behavior for signals recorded at the same electrode.

3.2.1 Effects of occluder placement on single TE neurons

The aim of the neurophysiological experiments was to establish how occlusion of different image parts influences responses of TE neurons. The previously described behavioral experiments successfully established that information is distributed non-homogeneously across natural scenes for monkey observers. Monkeys rely more on some regions than others to identify an image. These results were taken into account when constructing occluded versions of an image. Two categories of occluded stimuli were generated for each image. In the first category (the diagnostic conditions), occluders were placed so that the image regions with behavioral relevance remained visible. In the non-diagnostic conditions, these regions were covered and behaviorally irrelevant image regions were visible. Thus, the diagnosticity of different image regions determined the position and the shape of the occluders. It was in addition necessary to specify the occluders' extent. The experiments by Kovács et al. (1995) suggest that occluder size has an impact on the responses of TE neurons. Consequently, the extent of the occluder (the "visible stimulus size") was varied in addition to the occluder position. Occluders covered either 10, 30, or 50% of the images. To summarize, occluded image versions varied along two factors, diagnosticity and visible stimulus size.

Combining the two factors resulted in six occluded versions per tested image. Section 3.1.2 outlined the computation of these occluded stimuli. Since masks were based on each monkey's individual Bubbles results, the monkeys had their own, individual stimulus sets. Each monkey had behaviorally been tested on six natural scenes. To reduce the number of conditions, and also because no diagnostic regions could be obtained for some of the images, four out of the six images were selected per monkey, and used for further testing. These were scenes N1, N3, N5, and N7 for G00, and N2, N3, N5, and N6 for B98.

As described above, the effects of occlusion on TE responses were studied with respect to two parameters, diagnosticity and visible stimulus size. The behavioral effect of these parameters has already been described in Section 3.1.2. Most im-

portantly, the tests confirmed that occluder placement has a pronounced effect on the recognition rates for partially occluded images. Furthermore, the behavioral tests indicate that the visible stimulus size does not strongly influence the behavior. Even at the largest visible stimulus size, non-diagnostic conditions could not be identified. Similarly, the smallest visible stimulus size was sufficient to recognize images in the diagnostic conditions. Occluder placement and occluder extent thus have behaviorally different consequences. This opens up the possibility to study whether the influence of an occlusion parameter on behavior determines how much this parameter affects neural responses in TE.

Masking the images generated stimulus conditions with a known diagnosticity. At the same time, the occlusion introduced differences in the physical image properties between the stimuli. Since it would be difficult to disentangle the influences of occlusion and physical image properties on the neural responses, care was taken to equalize the low-level stimulus properties as much as possible. It is impossible to change the local properties of an image without affecting the image content, and possibly changing the diagnosticity of a region. Only global image characteristics were therefore adjusted to equalize the images. This concerned the mean luminance of the visible image regions, as well as the global contrast, measured as the standard deviation of the luminance of all visible image pixels. Mean luminance was set equal for all stimuli and all conditions; the global contrast was equalized for all seven conditions of an image (the full image and six masked versions).

Differences in local image properties between diagnostic and non-diagnostic stimuli were quantified thereafter. Every masked stimulus was filtered at three spatial resolutions to compute the local luminance and edge energy content, as has been described for the full images in Section 3.1.5. As before, edge energy was quantified for four different orientations. Only three spatial resolutions were considered, since the small visible image portion in some of the masked stimuli did not justify filtering at the coarsest spatial resolution. Luminance and edge maps were again combined in a luminance saliency, and an edge saliency map. Finally, the physical properties of each stimulus were summarized by averaging the computed feature values across all visible image parts. For the edge maps, the different orientations were also averaged. Stimulus regions covered by the mask were not taken into account when averaging, even if the filtering process generated differences in luminance or edge energy across these regions because of the spatial extent of the implemented filter kernels.

For each of the features, the 12 diagnostic stimuli in a monkey's stimulus set were compared to the 12 non-diagnostic stimuli by means of a t-test. In case of the stimulus set for monkey B98, the global adjustments were sufficient to remove any major local differences as well. Neither luminance nor edge energy values differed

at any of the three scales ($t(22) \leq 1.70$, $p \geq .10$ for the six tests). Similarly, there were no differences in either luminance saliency ($t(22) = .47$, $p = .65$) or edge saliency ($t(22) = -0.71$, $p = .49$). For monkey G00, stimuli were also equal in terms of their luminance content ($t(22) \leq .40$, $p \geq .70$ at the three spatial scales), and both saliency features (luminance saliency: $t(22) = 0.73$, $p = .47$; edge saliency: $t(22) = -1.09$, $p = .29$). However, edge energy showed differences between diagnostic and non-diagnostic conditions at two of the spatial resolutions. There were no significant differences at the finest resolution ($t(22) = 1.10$, $p = .28$); but at both coarser resolutions, there were more edges in the diagnostic than in the non-diagnostic stimuli ($t(22) \geq 3.07$, $p \leq .006$ for the two tests). Since this was the only noticeable difference between diagnostic and non-diagnostic stimuli, it was unlikely that observed neural effects could be traced back to low-level differences. This is even more the case because the difference was only present in the stimulus set of one of the monkeys. For any effects present in both monkeys, low-level differences could therefore be excluded as a reason.

The activity of single neurons in area TE was recorded while the monkeys performed a simple fixation task. Each trial began with the presentation of a fixation spot in the center of the screen. After the monkeys had acquired and maintained fixation on this position for at least 400 ms, one of the stimuli appeared behind the fixation spot for 500 ms. The monkeys were required to maintain the fixation during this interval. For successful fixation, they were rewarded with a drop of juice after stimulus offset. All 28 stimuli were shown in pseudo-random order, until between 10 and 20 repetitions had been collected per stimulus.

The behavior of an exemplar neuron recorded from monkey B98 is shown in Figure 3.20. This neuron responded to the presentation of scene N6 with an elevation in its firing rate. As can be seen in the raster plot in Figure 3.20B, presentation of the full – i.e. the unoccluded – stimulus elicited a vigorous response from the neuron about 100 ms after stimulus onset. The difference between the baseline firing rate computed from a 200 ms interval before stimulus onset, and the stimulus evoked firing rate in the interval from 100 to 400 ms after stimulus onset, yielded on average 4.9 spikes/s (see the black bar in Figure 3.20D). The significance of any response was tested by comparing the firing rate during stimulus presentation to the one in the baseline period with a t-test. Here, this difference in firing rate was significant ($t(20) = -2.80$, $p = .01$).

For the exemplar neuron, occluding parts of the image led to a general decrease in the firing rate. The reduction in firing rate depended on how much of the image remained visible; the more of an image was occluded, the stronger the firing rate was reduced. This trend can clearly be seen in the rastergrams shown in Figure 3.20C, as well as

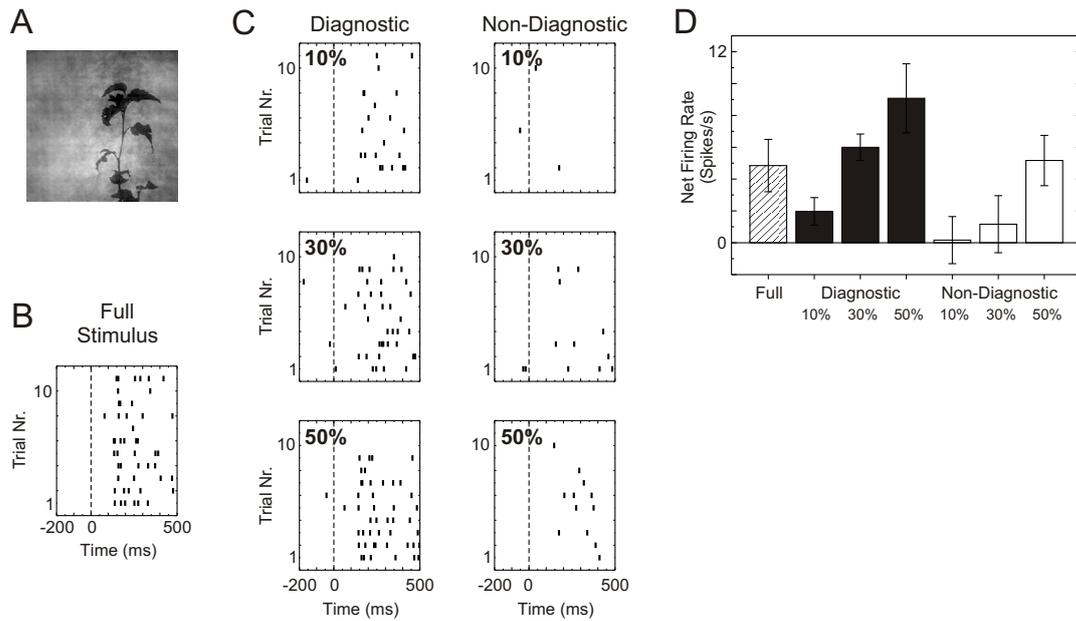


Figure 3.20: Behavior of an exemplar TE neuron recorded from monkey B98. A, Stimulus evoking the illustrated responses. B, Rastergram showing the response to the unoccluded (full) stimulus. Black bars in the rastergrams denote the occurrence of a spike at a specific point in time during the trial. Stimulus onset is at 0 ms. Along the y-axis, activity during repeated presentations of the same stimulus is plotted. C, Neural activity for occluded stimuli. The left column contains the diagnostic, the right column the non-diagnostic conditions. Labels indicate the percentage of the original stimulus visible through the mask. D, Net firing rate of the neuron for the different conditions. Net firing rate was computed by subtracting the baseline firing rate observed in the 200 ms interval prior to stimulus onset from the stimulus elicited firing rate during the interval from 100 to 400 ms after stimulus onset. In this and all following plots, the error bars indicate the standard error of the mean.

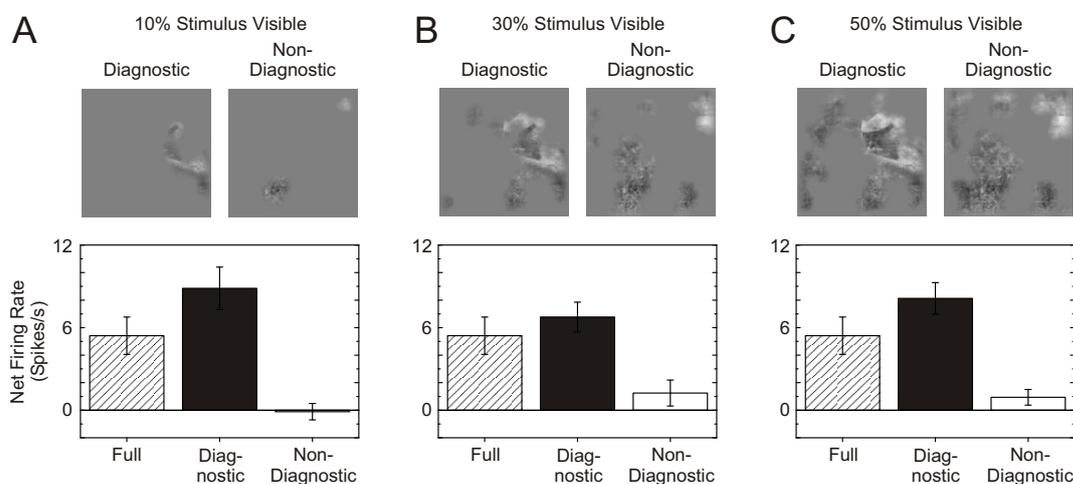


Figure 3.21: Another exemplar TE neuron, recorded from monkey G00. The neuron responded significantly to scene N2. A, Net firing rate for stimuli in which 10% of the original stimulus were exposed. The upper row shows the stimuli, the lower row the net firing rates. B, 30% of stimulus visible. C, 50% of stimulus visible. The net firing rate for the full stimulus is repeated in each bar graph to help the comparison between firing rates.

in the net firing rates plotted in Figure 3.20D. In addition to this general influence of occlusion, there was a pronounced difference between diagnostic and non-diagnostic conditions. In the diagnostic conditions, presentation even of the stimulus containing only 10% of the original stimulus was sufficient to elicit a visual response from the neuron that was significantly greater than baseline ($t(20)=-2.16$, $p=.04$). In contrast, when the diagnostic stimulus regions were occluded, 50% of the image had to be visible before the neuron started to respond significantly above baseline (non-diagnostic conditions, 10% of stimulus visible: $t(18)=-.011$, $p=.91$; 30% visible: $t(18)=-0.67$, $p=.51$, 50% visible: $t(18)=-3.45$, $p=.003$). As a result, firing rates in all diagnostic conditions were significantly greater than in the corresponding non-diagnostic conditions. The significance of the influence of both factors was confirmed by subjecting the net firing rates in the occluded conditions to an ANOVA with factors diagnosticity and visible stimulus size. With respect to the factor size, the ANOVA resulted in a significant main effect ($F(2,54)=8.04$, $p<.000$). In addition, there was a significant main effect for diagnosticity ($F(1,54)=5.02$, $p=.029$), but no significant interaction between diagnosticity and visible image size ($F(12,54)=1.59$, $p=.21$).

The behavior of two more neurons is illustrated in Figures 3.21 and 3.22 to exemplify the range of effects observed in the population of neurons. The neuron depicted in

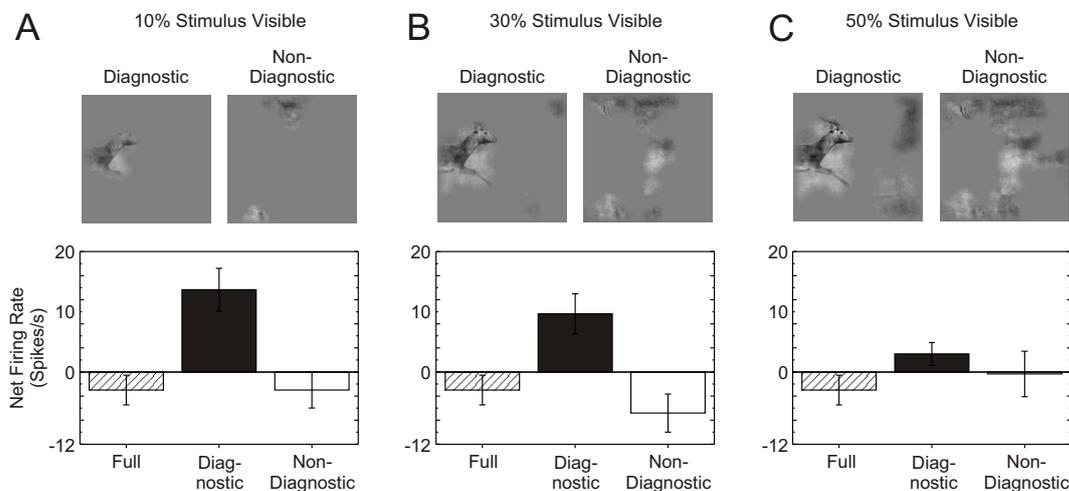


Figure 3.22: Third exemplar TE neuron, recorded from monkey G00. The neuron exhibited a very nonlinear behavior. Responses to scene N3 are plotted in the same format as Figure 3.21.

Figure 3.21 distinguished strongly between diagnostic and non-diagnostic conditions. It responded significantly to scene N3 ($t(30)=-3.99$, $p<.000$), and maintained its firing level for all three diagnostic conditions (t-test between the net firing rate for the full and each diagnostic condition; 10% of stimulus visible: $t(30)=-1.87$, $p=.07$; 30% visible: $t(30)=-1.30$, $p=.20$; 50% visible: $t(30)=-1.86$, $p=.07$). The neuron however did not respond to any of the non-diagnostic conditions above baseline level (non-diagnostic conditions, 10% of stimulus visible: $t(30)=0.17$, $p=.87$, 30% visible: $t(30)=-1.46$, $p=.16$, 50% visible: $t(30)=-1.80$, $p=.08$). Consequently, the difference between corresponding diagnostic and non-diagnostic conditions was significant for all three sizes. Again, an ANOVA was performed on the net firing rates in all six occluded conditions, using the factors diagnosticity and visible stimulus size. The only significant main effect was an effect of diagnosticity ($F(1,90)=82.52$, $p<.000$). Neither the effect of visible stimulus size ($F(2,90)=0.03$, $p=.97$), nor the interaction between the two factors ($F(2,90)=0.85$, $p=.43$) became significant.

The third neuron exhibited a nonlinear behavior. It did not respond to the full image (difference between baseline and stimulus evoked firing rate: $t(38)=1.29$, $p=.20$). However, when large parts of the image were occluded so that only diagnostic image regions were visible, the neuron vigorously responded to the stimulus (mean net firing rate for the condition in which 10% of stimulus were visible: 13.67 spikes/s; comparison to baseline level: $t(38)=-4.29$, $p<.000$). Increasing the exposed amount of the diagnostic region decreased the firing rate (mean net firing rate for the 30%

condition: 9.67 spikes/s, comparison to baseline: $t(38)=-2.35$, $p=.02$); the neuron became unresponsive when 50% of the stimulus were visible (mean net firing rate: 3 spikes/s; $t(38)=-1.38$, $p=.18$). In the non-diagnostic conditions, the neuron was mostly unresponsive, if not weakly inhibited (comparison to baseline: 10% of stimulus visible: $t(38)=0.86$, $p=.40$; 30% visible: $t(38)=2.08$, $p=.04$; 50% visible: $t(38)=0.09$, $p=.93$). The neuron's behavior again demonstrated differences between diagnostic and non-diagnostic conditions. More interestingly, the effects suggest a complex interplay between different scene regions in determining the neural response.

A total of 423 neurons (216 from monkey B98, 207 from monkey G00) were recorded from area TE. Since the results of both monkeys were similar, their data were pooled. From these neurons, only neurons were considered further which responded significantly and excitatory to at least one stimulus in one condition ($p<.05$ in the t-test between baseline and stimulus firing rate, Bonferoni corrected for the 28 necessary comparisons). Using this criterion, 129 neurons were included in the following analyses. Influences of occlusion on the neural population were quantified in two ways: First, to assess the dynamic behavior of the population, each neuron's response was characterized by a spike density function. Spike density functions were generated by convolving the spike trains of individual trials with a Gaussian kernel, and summing across trials. The spike density function represents the probability of a spike occurring at a certain time point after stimulus onset. Spike density functions were first computed for all 28 stimuli, and the data from different neurons were averaged. Differences in spike rates influence the amplitude of the spike density functions. When pooling data across neurons, the average would therefore be dominated by the neurons with high firing rates. Spike density functions were thus normalized before averaging. The comparisons of interested involved occluded conditions of one image, but not conditions of different images. Hence, the normalization was restricted to the conditions of a single image. Spike density functions were normalized to the maximum occurring across all seven conditions of an image. This normalization preserved the ratio between peak values from different conditions for each image. The normalized spike density functions were finally averaged across neurons and images.

Neurons most often only responded to one or two of the four images. An additional exclusion criterion was therefore used when averaging across images: From the complete data set collected for a neuron, the seven conditions of an image were only included in the average if the neuron responded significantly and excitatory to at least one of them (using the same p-level as above). As a consequence, each neuron could contribute in principle between 1 and 4 times to the population spike density function. Throughout the rest of the text, the term "case" will be used to denote

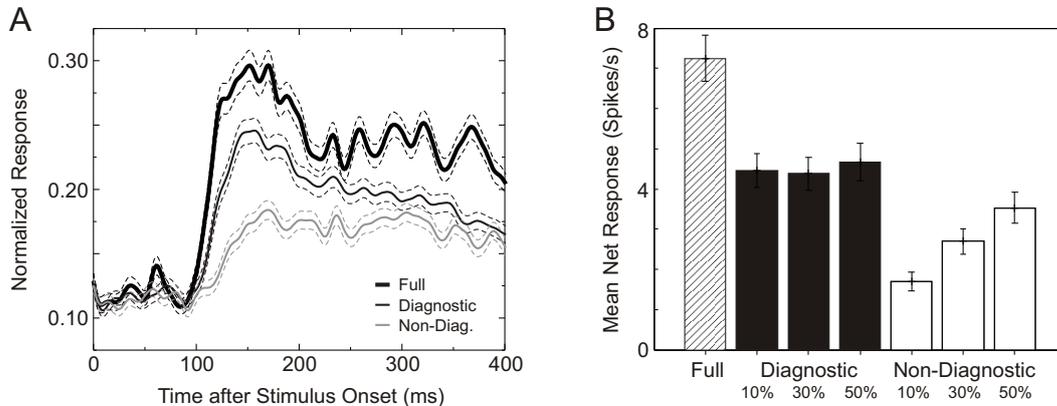


Figure 3.23: Behavior of the population of visually responsive TE neurons. *A*, Population spike density function as a function of time after stimulus onset. The three diagnostic conditions have been averaged, as well as the three non-diagnostic conditions. Solid lines show the average spike density function in a condition; the dashed lines correspond to the standard error of the mean at each time point. *B*, Average net response in each of the conditions. Both plots are based on the average of 220 cases generated from 129 neurons.

the data collected from a single neuron for all conditions of one image. With the chosen significance criterion, the average was based on 220 cases from the selected 129 neurons.

The population spike density functions for different conditions are plotted in Figure 3.23A. For easier interpretation, data from all diagnostic, and all non-diagnostic conditions have been averaged. All spike density functions showed a steep rise from baseline level at about 100 ms after stimulus onset. The peak level of the response was influenced by occlusion: In general, occlusion reduced the response rate. This effect was however much stronger for non-diagnostic than for diagnostic conditions. The differences between full, diagnostic, and non-diagnostic conditions were evident for the interval up to about 300 to 400 ms.

Based on these results, the time window from 100 to 400 ms was considered whenever the strength of a neural response to a stimulus was quantified. In the second analysis, the average net firing rates were computed as a function of the occlusion condition. Net firing rates were calculated as the average firing rate in the interval from 100 to 400 ms after stimulus onset, minus the firing rate in the baseline period 200 ms prior to stimulus onset. Net firing rates were averaged across neurons and images. As for the spike density functions, only the data from images evoking significant responses in at least one condition were included. The average net response was therefore based

on the same 220 cases. The average net response confirmed the observations made for the population spike density function (see Figure 3.23B). Occlusion generally reduced the average response rate. However, the response reduction was much stronger in the non-diagnostic conditions. Furthermore, the average showed little effect of the visible stimulus size on the response rates for diagnostic conditions. In contrast, the responses in non-diagnostic conditions were most strongly reduced when little of the stimulus was exposed by the mask. Response rate recovered in these conditions when more of the image became visible. This suggests that as long as diagnostic image material is visible, even very small image parts are sufficient to drive a neuron. In contrast, much more of an image has to be presented in the non-diagnostic conditions to evoke a response.

The general effect of occlusion on the neural responses was confirmed by performing a paired t-test between the unoccluded condition and any other condition. All six t-tests were highly significant ($t(219) \geq 4.43$, $p < .000$ in all cases). Systematic differences between the set of occluded conditions were assessed by performing a repeated measures ANOVA on the data (factors diagnosticity, visible stimulus size). The ANOVA resulted in a significant main effect of both factors (diagnosticity: $F(1,219) = 27.43$, $p < .000$; visible size: $F(2,438) = 7.22$, $p = .001$), as well as a significant interaction between them ($F(2,438) = 4.58$, $p = .01$). The interaction reflected the fact that diagnostic and non-diagnostic conditions were differently influenced by the visible stimulus size. Comparing diagnostic and non-diagnostic conditions separately at each size confirmed that the evoked firing rates were nonetheless significantly higher in diagnostic than non-diagnostic conditions, irrespective of the visible stimulus size (paired t-test, $t(219) \geq 2.17$, $p \leq .03$ for the three tests).

Differences between diagnostic and non-diagnostic conditions were not only evident at the population level. The influence of occlusion condition on individual neurons was tested by the following analysis. For each visually responsive case, the firing rates in the diagnostic conditions were plotted against the ones in the non-diagnostic conditions revealing the same amount of the stimulus. The same 220 cases as before were tested. Figure 3.24 shows the resulting scatter plots for the three sizes. Cases with a higher firing rate in a diagnostic than in a non-diagnostic condition – i.e. the cases displaying the same behavior as the general trend – fall above the diagonal in these plots, while cases with the reverse pattern fall below the diagonal. The number of cases belonging to these two categories was quantified for each visible stimulus size. Numbers are listed in Figure 3.24 above and below the diagonal in the upper right-hand corner of each graph. At all sizes, the proportion of cases in which diagnostic conditions evoked higher firing rates was significantly greater than the proportion of cases showing the reverse (χ^2 -test between the number of

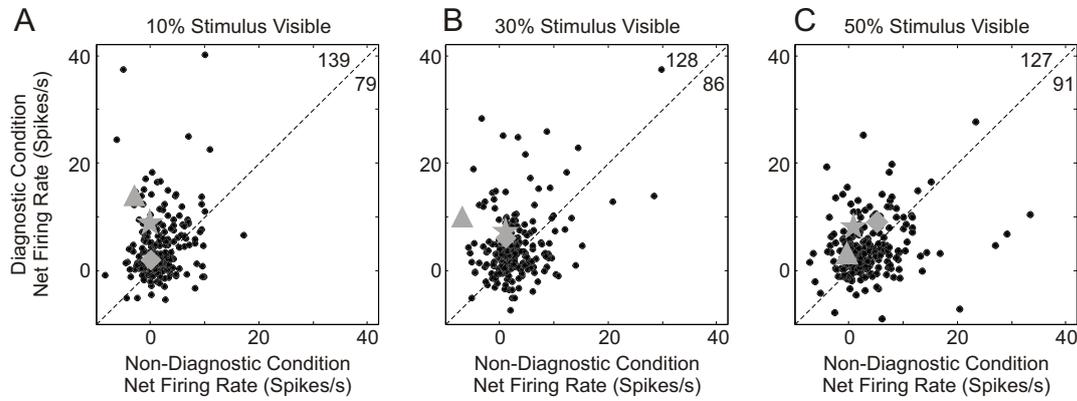


Figure 3.24: Influence of occlusion on neural responses for individual cases. For each case, the net firing rate in an diagnostic condition was plotted against the one in the corresponding non-diagnostic condition. Points above the diagonal indicate cases with a higher firing rate in the diagnostic condition. The numbers above and below the diagonal give the number of cases falling into the respective part of a graph. A, 10% of stimulus visible. B, 30% of stimulus visible. C, 50% of the stimulus visible. Gray symbols show the behavior of the exemplar neurons described above. \blacklozenge : exemplar unit 1; \blackstar : exemplar unit 2; \blacktriangle : exemplar unit 3.

cases in each category; 10% of stimulus visible: $\chi^2(1)=16.51$, $p<.000$; 30% visible: $\chi^2(1)=8.24$, $p=.004$; 50% visible: $\chi^2(1)=5.95$, $p=.02$). In conclusion, diagnostic conditions evoke higher firing rates in TE than non-diagnostic conditions.

When constructing masks for the occluded conditions, care was taken that the diagnostic and non-diagnostic conditions had no image regions in common. Inevitably, this made the visible stimulus regions in the two conditions cover different parts of the visual space. TE neurons have mostly been reported to have receptive fields larger than the size of the experimental stimuli with receptive field centers at the fovea; inhomogeneities in the receptive fields could nonetheless be present. It therefore had to be excluded that differences between diagnostic and non-diagnostic conditions were simply generated because of their spatial positions instead of their difference in behavioral relevance.

The influence of stimulation at different spatial positions was quantified for single neurons by the following analysis: Diagnostic and non-diagnostic conditions were constructed to show disparate image regions, and to expose up to 50% of the full image. Hence, the extent of the full image was completely probed across the six occluded conditions. Considering only the response to one image, a spatial response profile could therefore be computed for each neuron which specified the effect of

visual stimulation at a particular point in visual space. This spatial profile – the “response image” – was computed by determining the partially occluded versions of an image which elicited at least 70% of a neuron’s maximal response. The image regions visible in these conditions indicate at which spatial positions image material could evoke robust responses (the “effective” image regions). A criterion of 70% of the maximal response was chosen because it guaranteed that only conditions with robust responses were considered. However, similar results were obtained when including all conditions which evoked at least a firing rate of 50% of the maximal firing rate. Each response image showed the effective spatial locations for one of the images. Across the four images, diagnostic regions were usually located at different positions in space. Comparing a neuron’s four response images thus showed whether the neuron was mainly responsive to the presentation of image material at a particular location, or whether the effective spatial location depended on the image and followed the diagnosticity of image regions.

Response images for two exemplar neurons are shown in Figure 3.25. For both neurons, occluded conditions of three of the four images evoked responses at least as large as 70% of the maximal response. For each image, the regions visible in effective conditions are highlighted in Figure 3.25. The plots already indicate that for the different images, very different spatial regions evoke responses from the neuron. Responses to a partially occluded stimulus seem not to be generated because a certain spatial region was visible. Instead, the diagnosticity of an image region seems to trigger the responses.

Response images were computed for 46 out of the 129 visually responsive neurons. These neurons were selected because occluded versions of two or more images evoked responses at least as large as 70% of the maximal response. Across all response images, the regions effective in triggering a neural response spanned about 36.8% of the complete image extent. However, for individual neurons the overlap between the effective regions of different images only amounted to 10.3% of the image region. This suggests that in different images, very different regions elicited neural responses. In conclusion, neural responses were not bound to presentation of image material at a particular position. In contrast to a simple receptive field model, the diagnosticity determined whether image material presented at a particular position generated a neural response. Differences between diagnostic and non-diagnostic conditions are therefore not due to their different spatial positions.

The influence of two additional sources of artifacts needed to be considered. As explained at the beginning of the section, stimuli were equalized in terms of their global low-level image properties. With respect to the tested parameters, this seemed to be sufficient to also equalize more local properties of the images. However, although

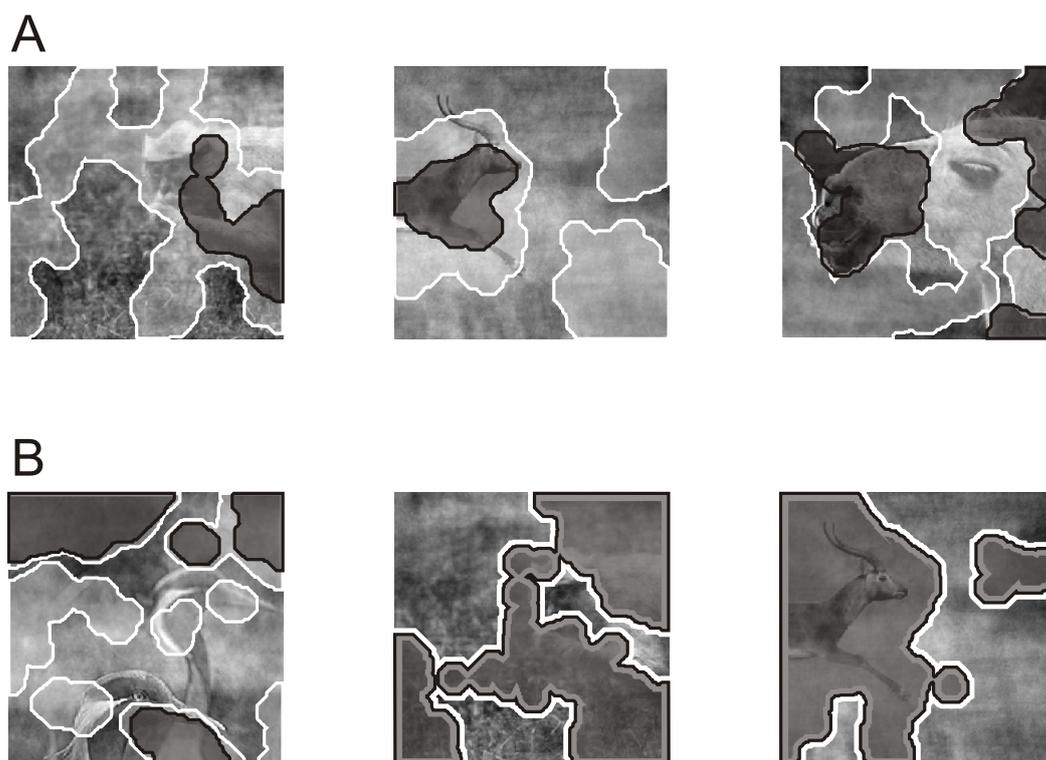


Figure 3.25: *Response images for two exemplar units. Black regions indicate which image regions were visible when the neuron fired at least with 70% of its maximal rate. The white regions enclose image regions visible in the largest diagnostic condition. A, Response images for a unit from monkey G00. B, Response images for a neuron from monkey B98. In the second and third plot, the response image completely overlapped with the diagnostic regions.*

the differences in the physical properties of diagnostic and non-diagnostic conditions were too small to be significant, they could have been potentiated in the neural responses. The effect of low-level physical image properties on the neural responses was therefore quantified in the following way. For each neuron, firing rates for a total of 28 stimuli were available. Each of these stimuli had slightly different low-level parameters, captured by the luminance, edge energy, luminance saliency and edge saliency maps computed before. For each stimulus, individual parameter maps were reduced to a single value by averaging across the visible stimulus parts. Again, edge energy maps of different orientations were averaged. In summary, each stimulus could be characterized by a total of eight parameters, corresponding to luminance and edge energy values at three spatial scales, and one value each for luminance and edge saliency.

To assess the influence of a parameter on the firing rate of a neuron, the 28 values of this parameter for the complete stimulus set were correlated against the firing rates for these stimuli. With eight different parameters, eight correlation coefficients were computed for each neuron. Taking only the visually responsive neurons into account, the number of neurons with significant correlations ($p < .05$) was counted for each of the parameters. For all eight parameters, these numbers were very low, ranging between six and 15 neurons out of the 129 neurons. No parameter seemed to influence firing rate more strongly than the others. In summary, since correlations were low between the computed low-level properties and the firing rate, differences at least in the captured physical properties could be excluded as a reason for the differences between diagnostic and non-diagnostic conditions.

Finally, TE neurons are known to be responsive to object parts. Diagnostic regions usually contained more object features than non-diagnostic regions. Since the presence of objects may be the reason for the diagnosticity of a region, there was no possibility to avoid this difference between conditions. However, the strategy of monkey B98 for some of the images made it possible to begin to disentangle the influence of object features and diagnosticity. Scene N2 was selected for further analysis. For this scene, diagnostic regions fell onto background regions. As a consequence, the non-diagnostic conditions contained more object features than the diagnostic conditions. If neural responses only follow the presence of object features, but not the diagnosticity, then the responses should be higher to the non-diagnostic versions of scene N2. In contrast, the encoding of diagnosticity predicts higher responses for the diagnostic conditions.

The analysis was restricted to the conditions with a visible stimulus size of 10%. At this occluder size, the differences in the amount of object features was maximal between conditions. A total of 19 neurons from monkey B98 were responsive to the

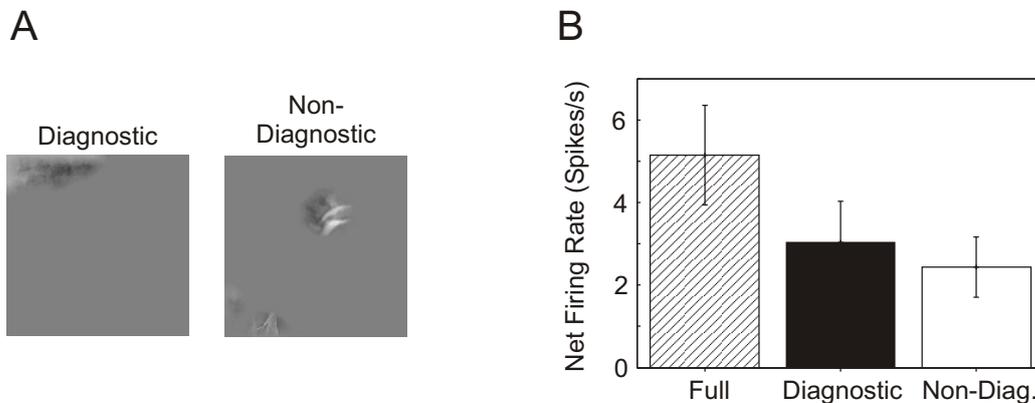


Figure 3.26: *Influence of physical image properties on neural responses. Neurons from monkey B98 responsive to scene N2 were selected (N=19). A, Diagnostic and non-diagnostic condition with a visible stimulus size of 10%. B, Mean response rate for the unoccluded, the 10% diagnostic, and the 10% non-diagnostic condition.*

scene. Their mean firing rate in the three relevant conditions is plotted in Figure 3.26. This plot shows that although here the non-diagnostic condition contained more object features, firing rates were not higher in the non-diagnostic condition. Instead, the trend was in the opposite direction. The difference between conditions was not significant for the mean net firing rate ($t(18)=0.47$, $p=.65$), but nonetheless more neurons (11 out of 19) had higher firing rates in the diagnostic than the non-diagnostic condition. The fact that firing rates were not higher for non-diagnostic than diagnostic conditions ruled out that differences in the presence of objects were the sole reason for the observed influences of diagnosticity on the neural firing rate. However, since for the selected neurons the firing rates did not distinguish between the two conditions, some influence of the presence of objects cannot be excluded.

In conclusion, neural firing rates were systematically affected by the diagnosticity of a stimulus. The higher the information content of a stimulus, the higher was the firing rate of the neurons. This effect was evident both on the population level, and for individual neurons. It could neither be attributed to stimulation at different spatial locations, nor to differences in low-level properties between diagnostic and non-diagnostic conditions.

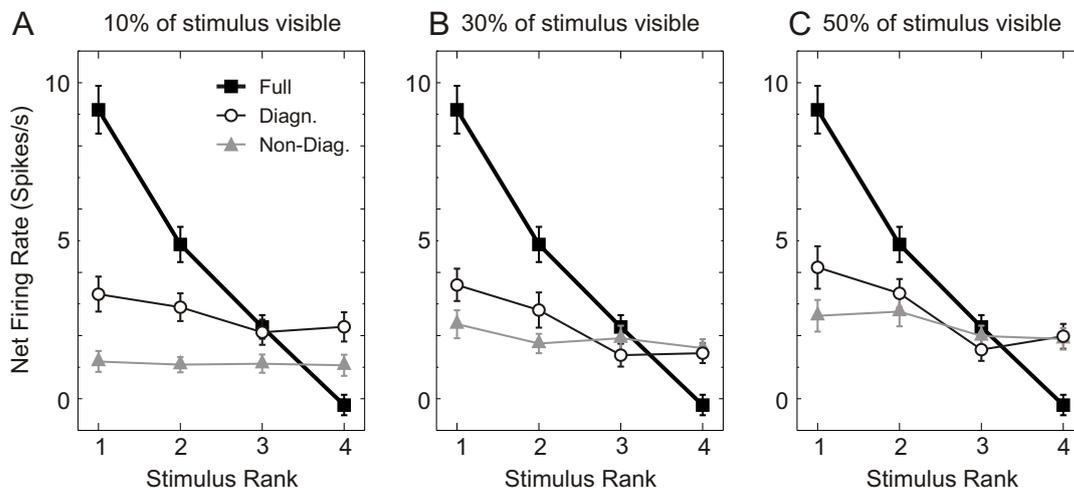


Figure 3.27: Tuning curves. Net firing rates are plotted as a function of stimulus rank, computed for the pool of 129 visually responsive neurons. The stimulus rank was determined from the responses in the full condition. 1 corresponds to the best, 4 to the worst stimulus. A, 10% of stimulus visible. B, 30% of stimulus visible. C, 50% of stimulus visible. The tuning curve in the full condition is identical in A-C.

3.2.2 Influence of occlusion on neural selectivity

Section 3.2.1 described the effects of occlusion on the responsiveness of TE neurons. In the analyses, results from different images were indiscriminately pooled. This section is concerned with the effects of occlusion on the selectivity of neurons, which is reflected in the spike rates evoked by different images. Every analysis was again based on the pool of the 129 visually responsive neurons selected by the criteria described in Section 3.2.1.

To assess the neural selectivity, a method from Kovács et al. (1995) was adopted. For each neuron, the four images were ranked according to their net firing rate in the unoccluded condition. Averaging across all visually responsive neurons, the net response was then calculated as a function of stimulus rank for each condition. The ranking was solely based on the responses in the unoccluded condition. Thus, if a neuron's stimulus preference changed for occluded stimuli, then there should be no relationship between the firing rate in the occluded condition and stimulus rank. In contrast, a monotonic decrease of the average response with stimulus rank in an occluded condition indicates that the stimulus preference was maintained despite the occlusion.

In the unoccluded condition, the pool of TE neurons exhibited a strong tuning to certain images (see Figure 3.27). While the neurons responded with an average of 9 spikes/s to their preferred image, the worst image did not evoke a response larger than baseline level. Occlusion changed the selectivity to a certain degree. Based on the previous analysis of the firing rates, an overall reduction in the firing rate had to be expected for the occluded conditions. However, for all occluded conditions, the firing rates for the best stimulus decreased, while the firing rates for the worst stimulus increased. This effect could only be generated because the stimulus preference of a neuron changed for occluded stimuli. The rank order was based on a neuron's behavior in the unoccluded condition. If with occlusion the neuron preferred a different image, then the firing rate for an image occupying one of the ranks 2 to 4 increased beyond the response to the image at rank 1. In turn, the differences in firing rates between ranks will be leveled out. Selectivity was changed for all conditions, but the effect seemed to be more pronounced for the non-diagnostic conditions. For these conditions, tuning curves remained almost flat irrespective of how much of an image was exposed. The degree of tuning was characterized by comparing the responses to the best and worst images with a paired t-test. For the non-diagnostic conditions, these effects did not reach significance at any of the visible stimulus sizes ($t(128) \leq 1.73$, $p \geq .09$). In contrast, the tuning curves in the diagnostic conditions recovered with increasing visible stimulus size. At a visible stimulus size of 10%, the best stimulus could not be distinguished from the worst based on the firing rates ($t(128) = 1.89$, $p = .06$). When 30% of an image were visible, both the best and the second best stimulus evoked significantly larger firing rates than the worst (best stimulus: $t(128) = 3.98$, $p < .000$; second best: $t(128) = 2.33$, $p = .02$). The same was the case for a visible stimulus size of 50% (best stimulus: $t(128) = 3.55$, $p < .000$; second best: $t(128) = 3.56$, $p < .000$).

Changes in selectivity were furthermore quantified by computing for each neuron the rank order of images in each of the occluded conditions. The rank of the best unoccluded image could then be determined in the occluded conditions. If a neuron's selectivity was preserved under occlusion conditions, then the best unoccluded image should always occupy rank 1 in the occluded conditions. If selectivity was however changed, the best unoccluded stimulus would occupy different ranks as well. Figure 3.28 shows the distributions of ranks for the best unoccluded image. As can be seen, although an image evoked the highest responses in the unoccluded condition, it did not necessarily do so in the occluded conditions. Instead, the best image from the unoccluded condition occupied all ranks. This was the case for all conditions. However, there were differences between the diagnostic and non-diagnostic conditions. These effects confirmed the conclusions drawn from the tuning curves. In the non-diagnostic conditions, the best unoccluded image was equally likely to be

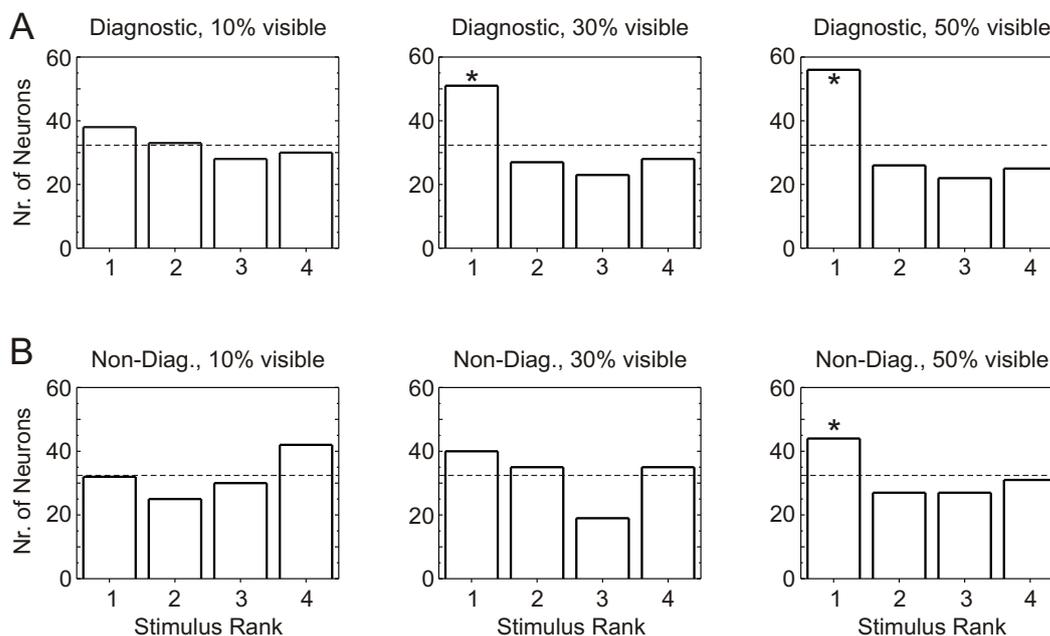


Figure 3.28: Agreement in selectivity between unoccluded and occluded condition. The best image was determined in the unoccluded condition. The rank of this image in the tuning curve in an occluded condition was then determined. Each plot shows the distribution of the ranks for one of the occluded conditions. The higher the percentage of rank 1 cases, the more agreement between the selectivity in the unoccluded and occluded condition. A, Diagnostic conditions. B, Non-diagnostic conditions. Dashed lines indicate the chance level, i.e. an equal number of neurons for all ranks. Stars denote rank 1 percentage significantly higher than chance (tested with a χ^2 -test).

found at all ranks. It was quantified for how many neurons the preferred stimulus remained the same despite occlusion. For all but the largest non-diagnostic condition, the percentage of these neurons did not significantly deviate from the percentage expected by chance (χ^2 -test between the number of rank 1 cases against chance level; visible stimulus size of 10%: $\chi^2=0.002$, $p=.96$; 30%: $\chi^2=1.86$, $p=.17$; 50%: $\chi^2=4.28$, $p=.04$). In contrast, the distributions for the diagnostic conditions show that here the selectivity indeed recovered. At least for the 30 and 50% condition, the number of neurons with matching selectivity in occluded and unoccluded conditions was significantly larger than chance (visible stimulus size of 10%: $\chi^2=1.03$, $p=.31$; 30%: $\chi^2=10.90$, $p<.000$; 50%: $\chi^2=17.49$, $p<.000$). In conclusion, occlusion did change the selectivity of the TE neurons. However, selectivity was changed more in the non-diagnostic than the diagnostic conditions.

If different images result in very different firing rates, observation of a specific response allows to determine which of the images was presented. The neural tuning curve therefore has an important functional implication, since it limits how much information a receiver – like a higher order brain area – can gain from the observation of a certain spike count about which image was presented in a trial. Using an information theoretic approach, the amount of information transmitted by a neuron can be quantified (Shannon & Weaver, 1949; Rieke et al., 1997). In this analysis, the brain is treated as a noisy communication channel. Images presented to the eye are the input into the communication channel. The output of the channel is taken to be the number of spikes generated by a TE neuron in a fixed time period. This spike count is used to deduce which image has been presented to the eye. In the given experimental situation, all four images are equiprobable as long as no spike count has been observed. After detection of a particular spike count, the uncertainty about the image identity can be reduced if the spike count is more likely to be elicited by certain images than by others. The reduction in uncertainty is equal to the mutual information between spike count and image identity. Computation of the mutual information involves the probability of showing a particular image, the probability of observing a certain spike count, and most importantly, the joint probability with which an image leads to a specific spike count. With four equiprobable images, as used in the present experiment, the mutual information could maximally reach 2 bits of information. At this optimal limit, the image identity could be determined without errors from the spike count. If less information is transmitted by a neuron, observation of a certain spike count reduces the number of possibly presented images, but does not allow a more precise determination of the stimulus identity.

In a first analysis, mutual information was computed from the spike counts in the interval from 100 to 400 ms after stimulus onset, the same interval used before for

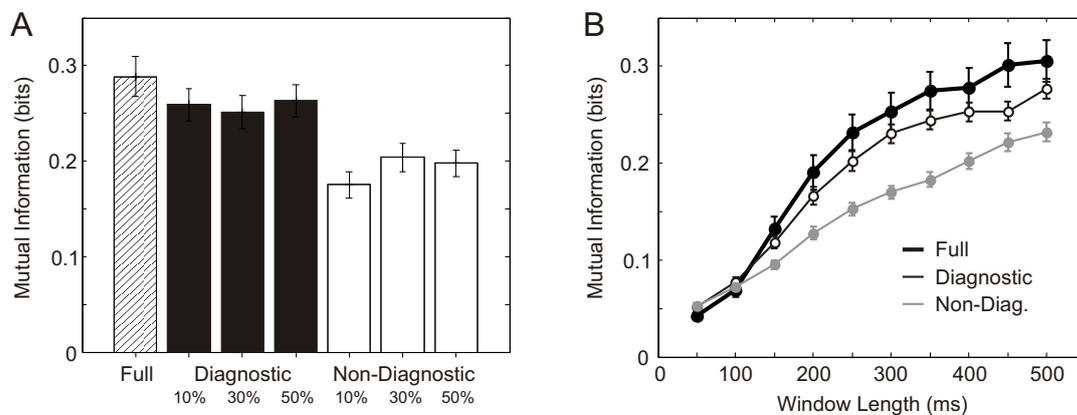


Figure 3.29: Influence of occlusion on the image information transmitted by the spike rate. *A*, Mutual information between the spike count and image identity in each of the seven conditions. Here, the spike count was obtained from the interval from 100 to 400 ms after stimulus onset. Data were averaged across 129 neurons. *B*, Influence of the counting window duration on the mutual information. Each window began at stimulus onset. Data from the diagnostic conditions have been averaged, as well as the data from non-diagnostic conditions.

computation of the net firing rates. Mutual information was computed separately for each condition and each neuron. The average over the pool of visually responsive neurons is plotted in Figure 3.29A. Mutual information was highest with unoccluded stimuli. Even in this condition, only about 0.3 bits of information were transmitted by the spike count, far less than the optimal value of 2 bits. In the diagnostic conditions, observation of a specific spike count conveyed as much information about the identity of the presented image as in the unoccluded condition. Performing paired t-tests between the mutual information in the full condition and the one in a diagnostic condition yielded non-significant results in all cases (10% visible: $t(128)=1.39$, $p=.17$; 30%: $t(128)=1.96$, $p=.05$; 50%: $t(128)=1.43$, $p=.16$).

In contrast, the spike rate transmitted significantly less information in the non-diagnostic conditions than in the unoccluded condition ($t(128)\geq 4.32$, $p<.000$ in all cases). In addition, the differences in mutual information between diagnostic and non-diagnostic condition were significant as well. Here, a repeated measures ANOVA was used to assess the robustness of the differences (factors diagnosticity, visible stimulus size). Mutual information was higher in diagnostic than non-diagnostic conditions (main effect factor diagnosticity: $F(1,128)=25.87$, $p<.000$). It was – in contrast to the average net firing rate – not influenced by the visible stimulus

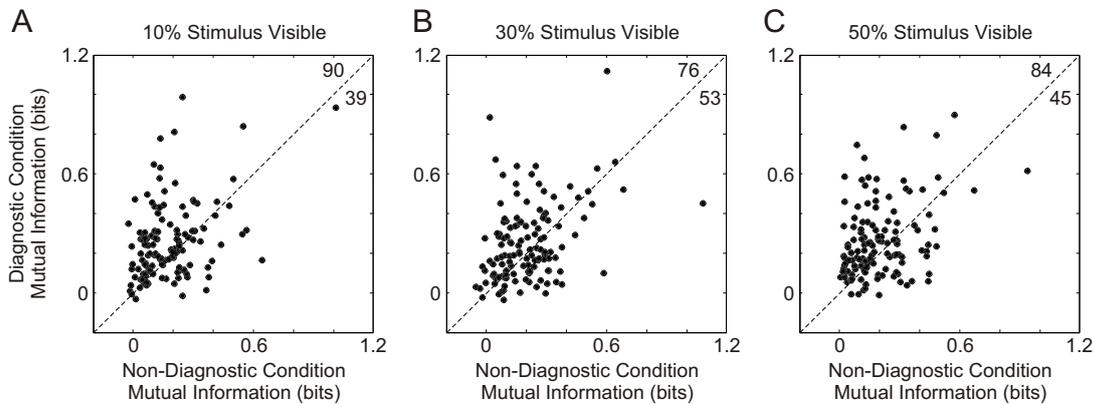


Figure 3.30: Differences between the mutual information in diagnostic and non-diagnostic conditions on the level of single neurons. For each neuron, the mutual information in an diagnostic condition is plotted against the one in the non-diagnostic condition. Numbers in the upper right-hand corner denote the number of units falling above or below the diagonal in a plot. A-C, Different amounts of visible stimulus size.

size (main effect factor size: $F(2,256)=.93$, $p=.40$; interaction between the factors: $F(2,256)=1.63$, $p=.20$).

Differences between the conditions were again not only present on the population level, but also for individual neurons. As for the net firing rate, a neuron's mutual information in a diagnostic condition was plotted against the one in the corresponding non-diagnostic condition (see Figure 3.30). Neurons were again divided into two categories, depending on whether the mutual information was higher in the diagnostic or the non-diagnostic condition. For all three visible stimulus sizes, the number of neurons in the first category was significantly higher than the number of neurons in the latter (χ^2 -test, $\chi^2 \geq 9.98$, $p \leq .002$ for all three tests).

To compute the mutual information, it was assumed that the spike rate of a neuron represents the output of a communication channel that transmits image identity. From the observation of a spike rate, the receiver tries to deduce which image has been presented to the monkey. The spike rate, however, depends on the interval during which action potentials are counted. So far, the counting interval was set from 100 to 400 ms after stimulus onset. In the final analysis, the start of the interval was fixed at the stimulus onset, while its length was varied from 50 to 400 ms in steps of 50 ms. This analysis made it possible to assess how a receiver starting to count spikes with stimulus onset would accumulate information with time, and to determine after which counting duration the diagnostic conditions began to deliver more information than the non-diagnostic ones.

Figure 3.29B shows the mutual information between spike rate and image identity as a function of counting window length. Data have been collapsed across visible stimulus size. Up to a window duration of 100 ms, i.e. for counting intervals ending before or at 100 ms after stimulus onset, mutual information was very low, and rose only slowly. However, since the spike density function showed an average neural response latency of about 100 ms, most neurons have not started to respond to the stimulus in this time period. Mutual information began to rise more steeply as soon as the counting interval ended after the neural response latency, i.e. at least 150 ms after stimulus onset. From this time point on, mutual information was consistently higher for diagnostic than for non-diagnostic conditions (t-tests at each window length: $t(772)=2.82$, $p<.005$ for all window lengths from 100 to 500 ms; only exception: window length 450 ms: $t(772)=2.39$, $p=.02$). For the full and the diagnostic condition, increasing the window duration increased the mutual information until an asymptote was reached at a window length of about 300 ms. In the non-diagnostic condition, mutual information seemed to be still increasing even with very long counting times.

Both the analysis of the tuning curves, as well as the mutual information between spike rate and stimulus identity highlighted another distinction between diagnostic and non-diagnostic conditions on the neural level. Neural selectivity was higher for diagnostic conditions. Based on the firing rates for diagnostic stimuli, the presented image could be identified with less errors than in the non-diagnostic conditions.

3.2.3 Influence of recording location on the behavior of single neurons

Since across all sessions the recording sites spanned a relative large extent of area TE, it could be tested whether neurons throughout the covered region behaved homogeneously, or whether neurons with a particular behavior clustered in a subregion of TE. On every recording session, the location of each electrode tip was noted. Electrode locations were specified by their distance from the chamber's center along the anterior-posterior and the medial-lateral direction. In addition, the recording depth was measured as the distance from the electrode tip to the border between the STS and the upper bank of TE (i.e. the ventral end of the STS). Based on these coordinates, the influence of the different stimulus conditions could be analyzed as a function of recording location. As before, the analysis collapsed responses to different images, including only images for which at least one condition evoked significant responses. On a few occasions, the recording depth could not be unambiguously determined with respect to the sulcal boundaries; these cases were excluded from

the analysis. Therefore, the data set contained 216 cases, 87 from monkey B98 and 129 from monkey G00.

Plotting the topography of occlusion effects required a very compact description of the behavior of individual cases. To this end, the effectiveness of the two stimulus manipulations, i.e. changes of diagnosticity and changes of visible stimulus size, was characterized for each responsive case. An experimental factor is very effective if the firing rate follows changes in this factor to a large degree. Effective factors therefore introduce large modulations in the firing rate. This also implies that changes in an effective factor underly large proportions of a neuron's variance when considering responses across trials of different conditions. Ineffective factors do not modulate the firing rate, and therefore contribute only little to the response variability. Effectiveness of diagnosticity and visible stimulus size were therefore quantified in terms of their explained variance. In this analysis, the two stimulus manipulations were treated as acting independently on the neural responses. First, the variance in firing rate was computed across all trials for a responsive case. Trials were then grouped according to the different levels of one of the factors, i.e. either according to whether the presented stimulus was diagnostic or not, or according to how much of an image was visible. The mean firing rate within each group was computed, and used to derive the variance of the group means. Assume that a firing rate was completely determined by the particular level of a factor. In this case, all firing rates within a group would be equal, and the variance between groups would equal the total variance of the whole data set. If, however, the factor level had no influence at all on the firing rate, the average firing rate would be the same for all factor levels, and the variance between groups would be zero. To compute a normalized measure, the explained variance was finally given by the ratio of the variance between groups and the total variance (Bortz, 1993).

The explained variance is a measure that captures trial-to-trial variability. It was therefore based on the firing rates from individual trials. In a later section, the explained variances for individual neurons will be compared against the one for local field potential sites. The latter data could not be corrected for baseline influences. To allow for a fair comparison, the firing rates were therefore also not corrected for the baseline firing rate when computing the explained variances.

To illustrate the possible outcomes of this analysis, the variance explained by either diagnosticity or visible stimulus size was computed for the three exemplar units shown in Section 3.2.1. Unit 1 (see Figure 3.20D) discriminated between diagnostic and non-diagnostic conditions, but the visible stimulus size exerted an even larger effect on the firing rates, increasing responses irrespective of a stimulus' diagnosticity. Consequently, changes in diagnosticity explained about 5% of the total variance,

while changes in visible stimulus size amounted to about 20% of the total variance. In contrast, the firing rates of unit 2 (see Figure 3.21) were mainly determined by the diagnosticity of a stimulus, with little influence of the visible stimulus size. Here, the factor diagnosticity explained 47% of the variance, while the factor visible stimulus size only contributed about 0.05%. Finally, the last unit (see Figure 3.22) displayed a more complex behavior. Here, differences in firing rates between diagnostic and non-diagnostic stimuli contributed about 16% to the total variance, while about 7% could be explained by how much of a stimulus was visible.

Based on the explained variances, the influence of each of the two stimulus manipulations on a neuron could now be expressed in a single number. Before analyzing any influences of recording locations, the dependency between the two factors was first investigated. Explained variances for both factors were computed from the same pool of trials. The explained variances for the two factors thus depend on each other. Assuming that no additional noise is present, they have to sum to 100%, since in this case the influences of the stimulus manipulations are the only source of variability in the data. Consequently, the two factors together have to explain the complete variance on the firing rate. In the real data, additional noise is present, and the explained variances of diagnosticity and visible stimulus size will sum to less than 100%.

Two types of dependencies are conceivable. First, the effect of the two stimulus manipulations could be positively correlated. In this case, neurons in which changes in diagnosticity robustly influenced the firing rate should also show reliable effects of the visible stimulus size. Second, neurons could be either encoding diagnosticity or visible stimulus size. In this case, high values of explained variances for one factor should be coupled with low values for the other factor. The latter was the case: Most cases with high explained variance values for diagnosticity had low explained variances for visible stimulus size, and vice versa. Figure 3.31A and B plot the explained variances for diagnosticity versus the ones for size. Each plot also shows the 90th percentile of the explained variance data of the respective factor. In both plots, it is evident that – with the exception of a few cases – cases above the 90th percentile for one factor were below the 90th percentile for the other factor. In conclusion, for these cases observed changes in the firing rate indicated the levels of only one of the factors but not both. The data suggest that diagnosticity and visible stimulus size are independently encoded in TE neurons.

Since each case was influenced only by either diagnosticity or visible stimulus size, cases were categorized as “diagnosticity” or “size” cases, based on whether their explained variance values fell above or below the 90th percentile for the corresponding factor. The recording location of these cases was then plotted. The center coor-

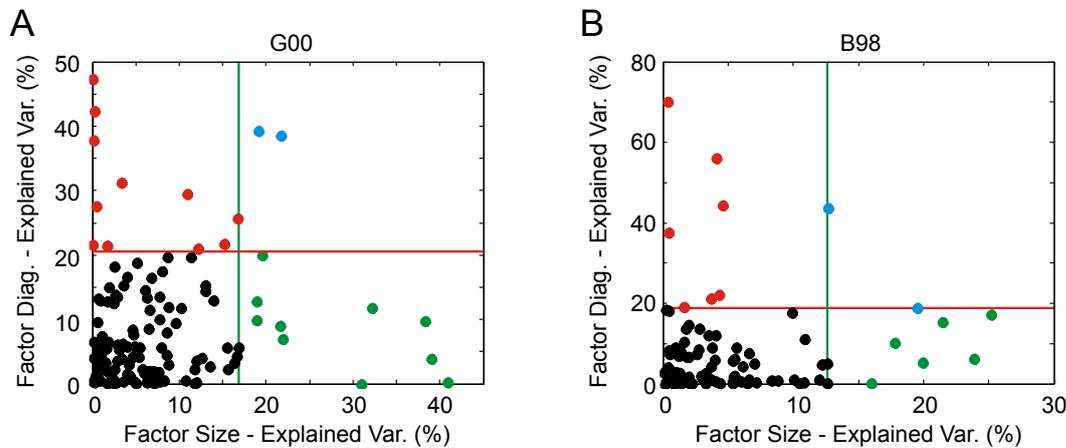


Figure 3.31: *Influence of the two stimulus manipulations. The variance explained by changes in diagnosticity is plotted against the variance explained by changes in the visible stimulus size. A, Results for monkey G00. B, Results for monkey B98. The green and red lines indicate the 90th percentile of the explained variances for the two factors. Red dots highlight cases above the 90th percentile for diagnosticity, green dots for visible stimulus size. Blue indicates cases which fall above the 90th percentile of both factors.*

dinates of the recording chambers were different for the two monkeys, and their data were thus treated separately. The resulting topographies are plotted as 2-D projections on coronal and sagittal views of the monkey brain in Figure 3.32. It was tested whether cases of the same type clustered in certain regions by splitting the covered recording area into two equally large subregions along the anterior-posterior axis, as well as the medial-lateral axis. Along the dorsal-ventral axis, the recording sites were categorized as either falling into the lower bank of the STS or lateral TE, depending on whether they fell above or below the white matter region separating the two regions. The boundaries between individual subregions are indicated in Figure 3.32 by the dashed lines. The number of diagnosticity and size cases in each of the subregions was then counted. Table 3.2 lists these numbers, in addition to the results of a binomial test used to assess the significance of differences along the three directions.

For the diagnosticity cases, no clustering was found along the AP direction. This was the case for both monkeys. With respect to recording depth, diagnosticity cases seemed to be present more often in the lower bank of the STS than in lateral TE. Although the difference between the number of cases in the two regions reached significance only for monkey G00, the same trend was present in the topography for monkey B98. Finally, there was also a tendency for diagnosticity cases to cluster

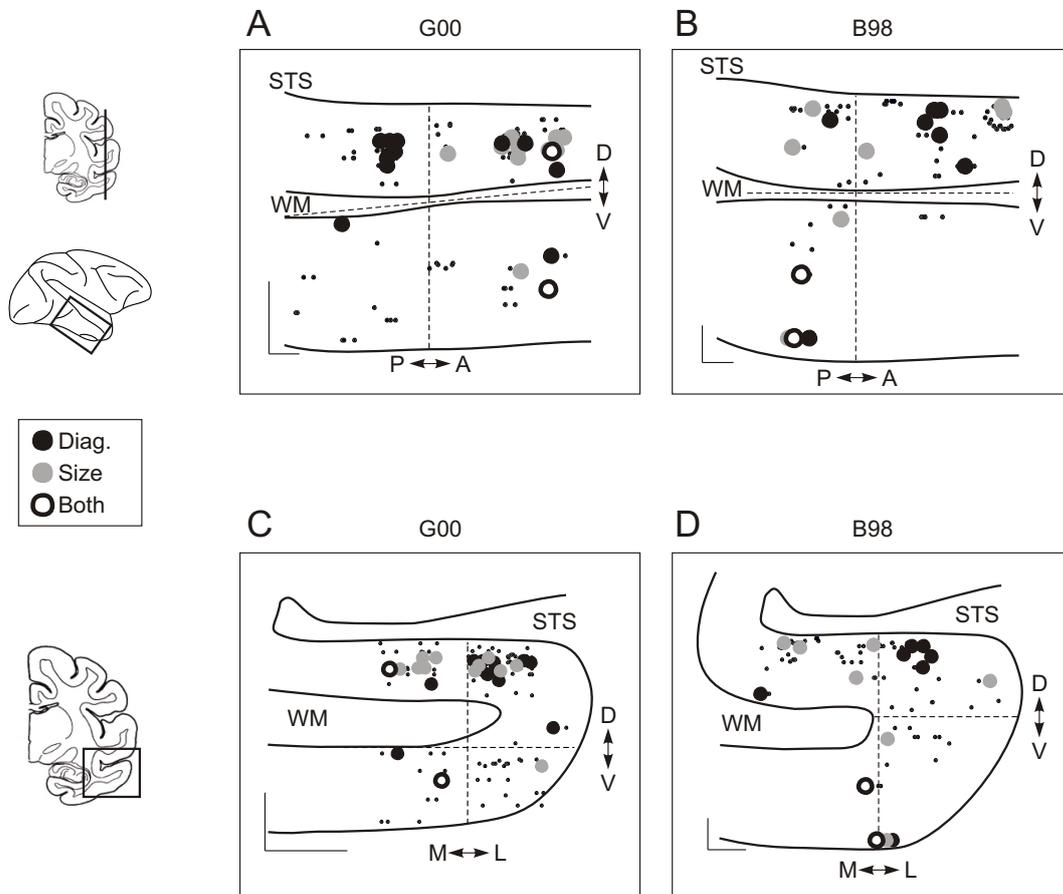


Figure 3.32: Location of diagnosticity and size cases, shown on 2-D projections of the monkey brain. A and C present data from monkey G00, B and D from monkey B98. A and B illustrate the distribution along the AP axis on a sagittal view of parts of the temporal lobe. This view is generated by cutting a slice along the line shown in the uppermost small figure on the left. For the boxed region in the second small figure on the left, the slice shows the lower bank of the STS separated from lateral TE by a band of white matter. C and D show the distribution along the ML axis on a coronal view of the region below the STS. The picture on the left indicates the position of this enlarged view. Each dot corresponds to the location of one case. Large dots show diagnosticity and size cases, while small black dots indicate the location of the rest of the cases not falling into these groups. To allow a better separation of different cases, the AP and ML positions of cases were randomly jittered by a small amount. Lines in the lower left corner of each graph are scale bars. The length of each scale bar leg is 1 mm. Dashed lines indicate the boundaries used to divide the recording region into smaller subregions. The views of the temporal lobes are not based on the exact anatomy of a monkey; instead they are only approximate sketches to help the interpretation of the plots. Abbreviations: WM - White matter; A - Anterior subregion; P - Posterior subregion; M - Medial subregion; L - Lateral subregion; D - Dorsal subregion; V - Ventral subregion.

Table 3.2: Number of diagnosticity and size cases falling into the different subregions of the covered TE part. For the subregions' labels, see Figure 3.32. The significance values give the p -values computed from a binomial test comparing the numbers in different subregions.

	G00		B98	
	Diagnosis- ticity	Size	Diagnosis- ticity	Size
A	6	13	5	3
P	7	0	4	6
Sig. A-P	1.0	<.000	1.0	.51
M	4	7	1	4
L	9	6	8	5
Sig. M-L	.27	1.0	.04	1.0
D	11	11	6	5
V	2	2	3	4
Sig. D-V	.02	.02	.51	1.0

in lateral positions, which was evident in a significant difference for the number of cases in the medial and lateral region in monkey B98, and a trend in the same direction for monkey G00. In monkey G00, size cases were clustered in the anterior part of the lower bank of the STS. However, this pattern was not consistent across monkeys, since the distribution of size cases was homogeneous in monkey B98.

So far, the locations of cases that could clearly be assigned as being influenced either by the diagnosticity of a stimulus, or its visible stimulus size, were plotted. To test for a systematic distribution present in the whole population, the explained variances of all cases were plotted against their AP position. Since the purpose of this analysis was to identify globally present relationships between the behavior of neurons and their location, the data from both monkeys were combined. The analysis was restricted to dependencies on the AP position because the recording sites in monkey G00 spanned a too small range of ML positions. The plots (see Figure 3.33) agreed with the above observations: No global pattern emerged for the distribution of diagnosticity sites along the AP axis. Although the explained variances for the factor diagnosticity were lowest for the most posterior locations, the range of values observed was relatively constant throughout the rest of the covered region. This was also reflected in the fact that the values from both monkeys were in good agreement. The independence of the explained variances from the AP position was confirmed by computing a correlation between the two parameters. Correlation coefficients were not significantly different from zero for the whole data set, as well as for each

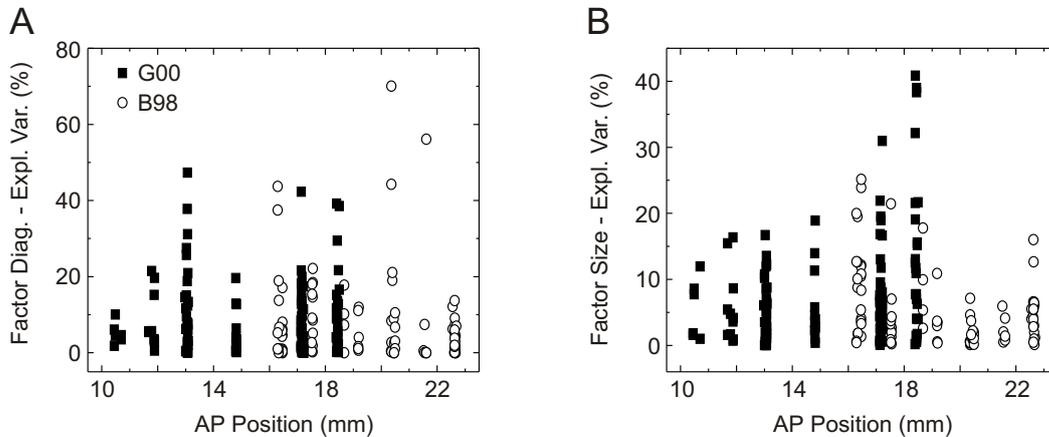


Figure 3.33: *Explained variance as a function of AP position. Data from both monkeys has been combined. A, Influences of diagnosticity. B, Influences of the visible stimulus size. Each dot represents data from one visually responsive case.*

monkey's data set tested independently ($p \geq .46$ for all three tests). For the variance explained by changes of the visible stimulus size, the results were incongruent between monkeys. While they increased with AP position in monkey G00, they decreased with AP position in monkey B98. As a result, the correlations between explained variance and AP position were significant when the monkeys were analyzed separately (G00: $r = .28$, $p = .001$; B98: $r = -.32$, $p < .000$), but not for the whole data set ($r = -.02$, $p = .77$).

In summary, neurons throughout the whole anterior-posterior extent of the covered TE region were influenced by changes in the diagnosticity of a stimulus. There was a tendency for these neurons to cluster more in the lateral part of the lower bank of the STS than in more medial and ventral parts. Neurons for which the firing rate was modulated with the visible image amount were distributed homogeneously throughout the whole recording region.

3.2.4 Representation of occluded stimuli by the LFP

The analysis so far was concerned with the behavior of single TE neurons. Their activity was measured by identifying action potentials in the signal recorded at an electrode, and computing the timing of these action potentials. Any other changes in the neural signal were discarded. However, the neural signal is indeed a comprehensive signal consisting of several components (for a thorough summary of the

topic, see Logothetis, 2002), and the action potentials analyzed so far are actually superimposed on relatively slow varying field potentials. To capture all signal components, spikes and slower voltage variations are usually separated from one another by applying high- and low-pass filtering, respectively, to the signal detected at an electrode. The high-pass filtered version of the signal contains the large action potentials generated by neurons very close to the electrode tip. It also reflects the weighted average of the spiking activity of neurons within a sphere of about 0.3 mm radius with the electrode at its center, a signal component called the multi-unit activity. The low-pass filtered version – the so called local field potential (LFP) – on the other hand is thought to predominantly represent synaptic events like the excitatory or inhibitory postsynaptic potentials. In the LFP, signals from the neural population within 0.5 - 3 mm of the electrode tip are captured.

This section describes the changes induced in the LFP by the different conditions. During the recordings, the LFP was isolated by applying a low-pass filter (1 - 100 Hz) to the detected signal. To remove any slow drifts and other artifacts, the recorded LFP signals were additionally filtered offline between 5 and 80 Hz. Because depending on the signal quality different amplification gains had been used during the recordings, any LFP data needed to be normalized before signals from different sites or sessions could be combined. For this purpose, the 100 ms baseline period preceding stimulus onset was chosen as a reference. In each trial, the mean and standard deviation of the LFP signal were computed for the baseline period, and the LFP signal from the whole trial was then z-transformed using these parameters. As a consequence, the LFP from each trial was centered around zero in the baseline period, and was given in units of baseline standard deviation.

Fluctuations in the LFP traces from individual trials were usually large, and any influences of the stimulus were masked by noise. Systematic influences of visual stimulation on the LFP were therefore first identified by averaging the LFP from trials with the same stimulus, computing the so called visual evoked potential (VEP). Before averaging, all trials were aligned with respect to the stimulus onset. Using this procedure, consistent, stimulus locked LFP components were enhanced in the average, while stimulus uncorrelated, inconsistent components were suppressed.

Besides assessing stimulus influences, the VEP was also used to determine whether a site was visually responsive to a stimulus at all. To be included in any of the further analyses, the VEP amplitude in at least one condition of a particular image needed to exceed either 1.5 or -1.5 times the baseline standard deviation at three consecutive time points during the stimulus presentation. Different thresholds were tested and led to comparable results. When collapsing data across images, the responsiveness of a site was tested separately for each image. As for the single units, responses

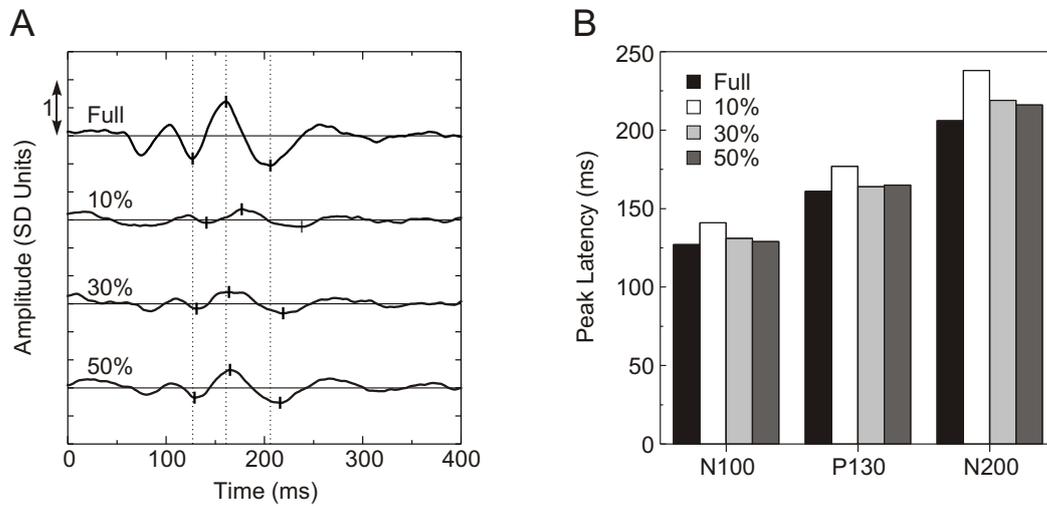


Figure 3.34: Identification of VEP components. *A*, Grand average VEPs from monkey B98. VEPs were averaged across images and recording sites. Diagnostic and non-diagnostic conditions of the same were pooled; conditions are referenced by the amount of image visible. The dashed lines indicate the latency of the three prominent peaks in the VEP for the unmasked images. Peak latencies in other conditions are indicated by the small markers. 0 ms corresponds to stimulus onset. *B*, Peak latency of the three components, compared across conditions.

in the seven conditions for an image were included if at least one condition evoked large enough VEP responses. A single site could therefore contribute multiple times to the average.

Applying this selection criterion, VEPs of the different conditions were pooled across images and recording sites. For monkey G00, 293 cases from 102 recording sites contributed to the analysis; for monkey B98, averages were based on 165 cases from 112 sites. On first inspection, a number of peaks could be identified in these grand averages. Peak amplitude seemed to be influenced by stimulus condition, but there were also differences in peak latencies between conditions, as well as between monkeys. In turn, before the peak amplitude could be analyzed further, the latencies of interesting peaks needed to be determined. Since the main factor influencing latency seemed to be the visible stimulus size, and non-diagnostic conditions did sometimes not lead to identifiable peaks in the grand averages, peak latencies were determined from VEPs in which diagnostic and non-diagnostic conditions with the same visible stimulus size were collapsed. The analysis was performed separately for each monkey.

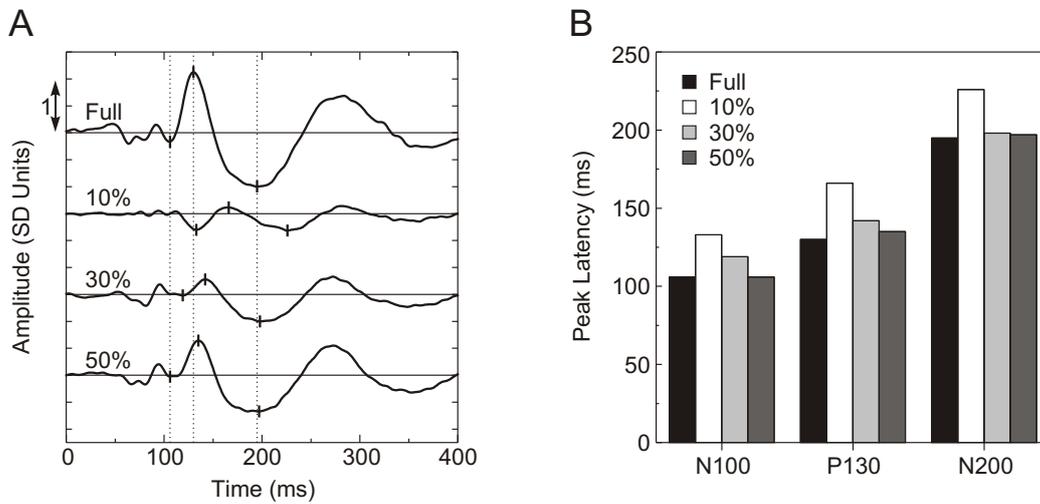


Figure 3.35: VEP components for monkey G00. A, Grand average VEPs. B, Latency of the VEP peaks for the different conditions. Layout of the graph is identical to Figure 3.34.

Figure 3.34 and 3.35 plot the resulting grand average VEPs for the two monkeys. In each of the conditions, the latency of three peaks was identified. These were a negative peak around 100 ms after stimulus onset, a positive peak around 130 ms, and a negative peak around 200 ms. The analysis was restricted to these three peaks, since they were detectable in both monkeys. Following a convention used in the EEG literature, the three VEP components will be referred to as N100, P130, and N200. Peak latency of a component was determined in the average VEPs as the time point of the maximal or minimal VEP amplitude. For both monkeys, peak latency was clearly influenced by how much of a stimulus was visible: When only 10% of a stimulus could be seen, peak latencies were longest; increasing the visible stimulus amount decreased the latencies almost to the same values as in the unoccluded condition. In addition, there was an overall difference between the two monkeys: Peaks appeared on average about 18 ms later in monkey B98 than in monkey G00. Both effects could be seen for each of the three peaks.

The goal of the subsequent analysis was to quantify differences between conditions in the mean VEP amplitude during the N100, P130, and N200. Because of the different peak latencies, different time windows had to be specified for each condition and monkey. Time windows were centered on the peak latencies determined from the grand average, and had a duration of 20 ms. Increasing the duration to 30 ms did however not change the results. Identical time windows were used for diagnostic

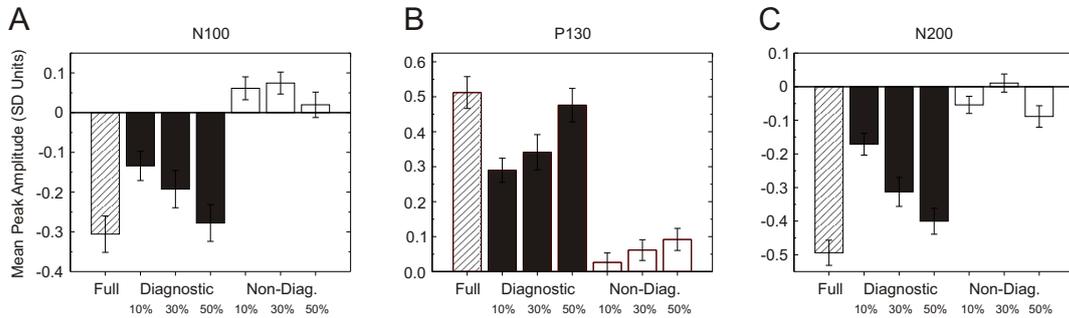


Figure 3.36: Mean VEP amplitude during the N100, P130, and N200 for monkey B98. VEP amplitudes were averaged in 20 ms time windows around the peak latencies for the different components. A-C, Mean VEP for the N100, P130, and N200.

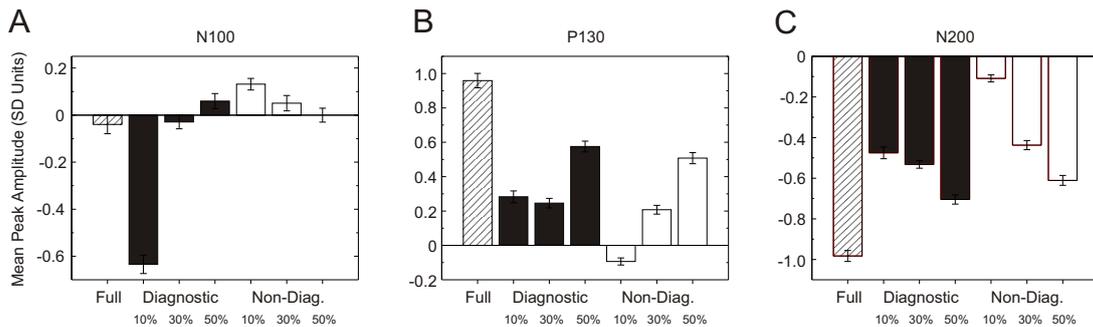


Figure 3.37: Influence of stimulus conditions on the mean LFP amplitude for monkey G00. A, Mean VEP in the N100 period. B-C, Mean responses for the P130 and N200.

and non-diagnostic conditions of the same visible stimulus size. Within each time window, the VEP amplitude was then averaged. The mean VEP amplitude was averaged across images and recording sites. Figures 3.36 and 3.37 show these averaged mean amplitudes for all seven conditions during the time interval of the N100, P130, and N200. Since for the LFP both monkeys led to different results, their data were analyzed separately.

For monkey B98, there was a clear-cut distinction between diagnostic and non-diagnostic conditions. While diagnostic conditions evoked VEP amplitudes of the same polarity and almost the same magnitude as the unoccluded condition, non-diagnostic VEPs had only very small amplitudes that sometimes even had different polarities than the VEP in the unoccluded condition. VEP amplitude was also modulated by how much of a stimulus could be seen through a mask: The more of a

stimulus became visible, the larger the amplitudes of the VEP. The significance of these observations was verified by performing a repeated measures ANOVA with factors diagnosticity and visible stimulus size on the VEP amplitudes. ANOVAs were computed for each of the three time intervals. For the time window corresponding to the N100, as well as the P130, both factors resulted in a significant main effect (N100: diagnosticity: $F(1,164)=44.30$, $p<.000$; size: $F(2,328)=5.61$, $p=.004$; P130: diagnosticity: $F(1,164)=55.18$, $p<.000$; size: $F(2,328)=10.79$, $p<.000$), while the interaction between the factors was not significant (N100: $F(2,328)=1.98$, $p=.14$; P130: $F(2,328)=2.14$, $p=.12$). For the N200, the main effects again were highly significant (diagnosticity: $F(1,164)=43.57$, $p<.000$, size: $F(2,328)=8.81$, $p<.000$), but this time the interaction also became significant ($F(2,328)=8.81$, $p<.000$). Performing paired t-tests between corresponding diagnostic and non-diagnostic conditions confirmed that they differed reliably, irrespective of how much of the stimulus was visible ($t(164)\leq-2.84$, $p\leq.005$).

For monkey G00, the effects of different stimulus conditions on the VEP amplitude were more complex. In general, while for B98 diagnosticity had been the main determinant of the VEP amplitude, both factors influenced the VEP of monkey G00. Also, differences between the different temporal VEP components were more pronounced for G00 than for B98. In the case of the N100, all conditions with the exception of the smallest diagnostic condition led to very small VEP amplitudes. For the P130 and the N200, diagnostic conditions evoked slightly larger VEP amplitudes than the non-diagnostic conditions. Differences between diagnostic and non-diagnostic conditions were most pronounced when only 10% of a stimulus were visible. For the diagnostic conditions, showing only 10% of a stimulus elicited the same responses as showing 30% of a stimulus. In contrast, if the visible stimulus regions were not diagnostic, showing 10% of an image led to small VEP amplitudes, in the case of the P130 even with the opposite polarity than in the unoccluded condition. Increasing the visible stimulus size increased the amplitude of both diagnostic and non-diagnostic conditions. These effects were more pronounced for the non-diagnostic conditions. As a consequence, VEP amplitudes at a visible stimulus size of 30 and 50% did no longer strongly distinguish between diagnostic and non-diagnostic conditions.

For each of the time intervals, differences between conditions were again tested using a repeated measures ANOVA with the same factors as above. For all three VEP components, both main effects, as well as the interactions became significant (diagnosticity: $F(1,292)\geq 42.58$, $p<.000$; size: $F(2,584)\geq 100.98$, $p\leq.000$; interaction: $F(2,584)\geq 33.90$, $p<.000$), making a comparison between the corresponding diagnostic and non-diagnostic conditions necessary. In the case of the N100, differences were

significant when 10 or 30% of a stimulus were exposed (paired t-tests between diagnostic and non-diagnostic conditions; 10%: $t(293)=-20.02$, $p<.000$; 30%: $t(293)=-2.33$, $p=.02$), but not when 50% of the stimulus were visible ($t(293)=1.67$, $p=.10$). For the P130, the only significant difference was obtained when comparing diagnostic and non-diagnostic conditions with 10% visible stimulus size (10%: $t(293)=9.68$, $p\leq.000$; 30 and 50%: $t(293)\leq 1.93$, $p\geq.051$). In contrast, for the N200 all diagnostic and non-diagnostic conditions differed significantly ($t(293)\leq -3.51$, $p\leq.001$).

To summarize, occlusion did have systematic effects on the LFP behavior. Computing the grand average VEPs showed very consistent differences between diagnostic and non-diagnostic conditions in monkey B98. For monkey G00, the diagnosticity and the visible stimulus size influenced the VEP equally strongly.

3.2.5 Influence of recording location on the LFP behavior

For the rest of the analysis, the influences of the different stimulus manipulations on the LFP were quantified by computing the variance in the LFP amplitudes that could be attributed to changes of either diagnosticity or stimulus size. As for the single units, the explained variances for each factor were computed at each recording site. Explained variances quantify how much of the variance occurring at a site could be attributed to modulations because of changes in one factor. The analysis therefore needed to be based on the data from individual trials. Besides the corrections already performed, there is currently no method available to correct the LFP data from individual trials for the ongoing activity. Thus, the raw LFP amplitudes from each trial were used to compute the explained variances. Explained variances were calculated separately for the N100, P130, and N200. As before, 20 ms time windows were constructed around the peak latency appropriate for a condition, and each trial's LFP amplitude was averaged over these time intervals. Using these mean LFP amplitudes, the same analysis could be carried out that had before been applied to the firing rate of single neurons (see Section 3.2.3). The analysis collapsed the LFP responses to different images, while only including responsive cases. In a few sessions, the recording depth could not be clearly identified with respect to the sulcal boundaries. These sessions were excluded. A pool of 278 cases from monkey G00 and 158 cases from monkey B98 was analyzed.

Similar to the analysis performed on the neural data, the explained variance data were first used to investigate whether separate pools of LFP sites responded to changes in the two factors, or whether the influences of the two manipulations on a LFP site were similar. The interdependence of the two factors was assessed by

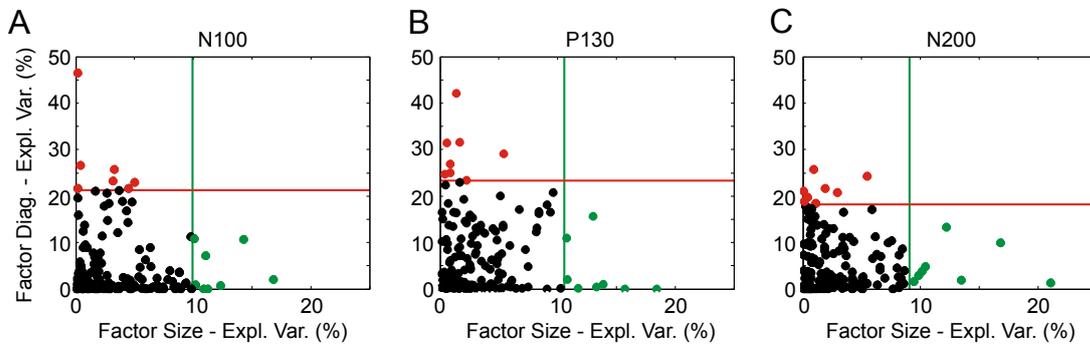


Figure 3.38: Dependency of the variance explained for diagnosticity on the variance explained for visible stimulus size. A-C correspond to the three LFP peaks. Plots pool the visually responsive LFP cases from monkey B98. Red and green lines indicate the 95th percentile of the data for one of the factors. Colored dots highlight cases falling above these percentiles.

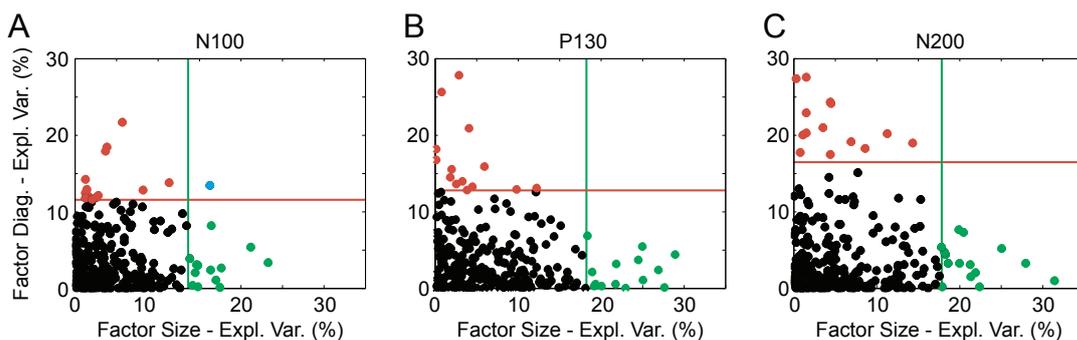


Figure 3.39: Variance explained for diagnosticity versus the variance explained for visible stimulus size for the LFP cases of monkey G00. A-C, Different VEP components. As in Figure 3.38, the colored lines indicate the 95th percentile for a factor. Red dots show cases above the 95th percentile for diagnosticity, green dots for visible stimulus size. Blue dots indicate cases above both percentiles.

plotting the variance explained by diagnosticity against the variance explained by the visible stimulus size (see Figure 3.38 and Figure 3.39). As for the single unit data, these plots demonstrated a negative correlation between the two factors. Cases where diagnosticity explained a large amount of the variance had low values for the variance explained by visible stimulus size, and vice versa. Cases could therefore be unambiguously characterized as either being influenced by diagnosticity, or by stimulus size.

The grand average VEPs discussed in Section 3.2.4 indicated a different LFP behavior in the two monkeys. The data suggested that while diagnosticity was the main factor influencing the LFP for monkey B98, both diagnosticity and visible stimulus size influenced the LFP amplitudes in monkey G00. However, judging from the explained variances, the LFP behavior seemed not so different in the monkeys. At least, recording sites could be identified in both monkeys where either diagnosticity or visible stimulus size strongly affected the LFP amplitudes. It was therefore of interest whether differences in the VEPs of the two monkeys would disappear if they were not computed across the whole population, but across selected sites. For each monkey, cases were selected where the two stimulus manipulations exerted strong effects. The cases were selected separately for each of the three VEP components. They had explained variance values above the 95th percentile for either diagnosticity (the diagnosticity sites), or the visible stimulus size (the size cases) for one of the VEP components. VEP amplitudes were then averaged across the diagnosticity or the size cases. Figures 3.40 and 3.41 plot the resulting average VEPs. It is clearly evident that the monkeys' VEPs now became more similar. Most importantly, the diagnosticity sites from both monkeys show prominent differences between diagnostic and non-diagnostic conditions. In contrast to the grand average data, even in monkey G00 the VEPs from the diagnosticity sites show almost no influence of the visible stimulus size. This was confirmed by performing a repeated measures ANOVA with factors diagnosticity and visible stimulus size on the VEP amplitudes for the diagnosticity cases of monkey G00. At the N100, both factors led to a significant main effect (diagnosticity: $F(1,14)=45.81$, $p<.000$; size: $F(2,28)=12.03$, $p<.000$). There was no significant interaction between the factors ($F(2,28)=2.30$, $p=.12$). For the P130, VEP amplitudes only distinguished between diagnostic and non-diagnostic conditions; there was no significant influence of the visible stimulus size (diagnosticity: $F(1,13)=160.14$, $p<.000$; size: $F(2,26)=.99$, $p=.39$; interaction: $F(2,26)=2.59$, $p=.09$). Finally, at the N200, both the main effect of diagnosticity, as well as the interaction between both factors were significant (diagnosticity: $F(1,13)=760.07$, $p<.000$; size: $F(2,26)=1.14$, $p=.33$; interaction: $F(2,26)=6.56$, $p=.005$). A paired t-test at each of the three visible stimulus sizes confirmed that despite the interac-

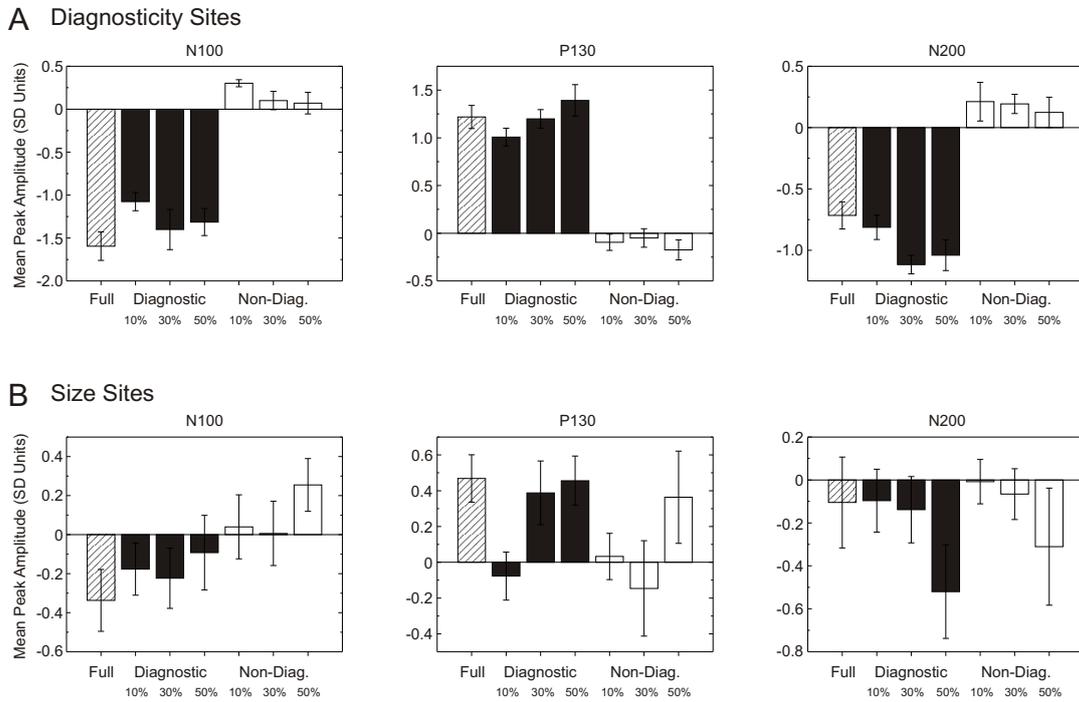


Figure 3.40: VEP amplitudes for selected LFP cases. LFP cases in monkey B98 were selected which either were strongly influenced by diagnosticity or the visible stimulus size. A, VEP amplitudes averaged over the diagnosticity cases for the N100, P130, and N200. B, Average VEP amplitudes for the size cases. VEPs of between 8 and 10 cases were averaged.

tion, VEP amplitudes were significantly larger in a diagnostic condition than in the matching non-diagnostic condition ($t(13) \leq -10.32$, $p < .000$ for all three tests).

Finally, the influence of recording location on the behavior of the LFP was assessed. Cases were again categorized as diagnosticity or size cases based on the explained variances. The analysis proceeded separately for each time window. The distribution of the selected cases along the AP axis is plotted in Figures 3.42 and 3.43. Since no consistent pattern emerged along the ML axis, the topographies along this direction are not plotted. Instead, the numbers of cases falling into the medial and the lateral part of the recording area are given in Table 3.3.

As a first result, Figures 3.42 and 3.43 indicate that at different time points, different populations of recording cases were strongly influenced by the stimulus manipulations. Whether these cases clustered in particular regions was tested by dividing the recording region into an anterior and posterior subregion, as well as a dorsal and ventral subregion. The same divisions as in Section 3.2.3 were used. The number

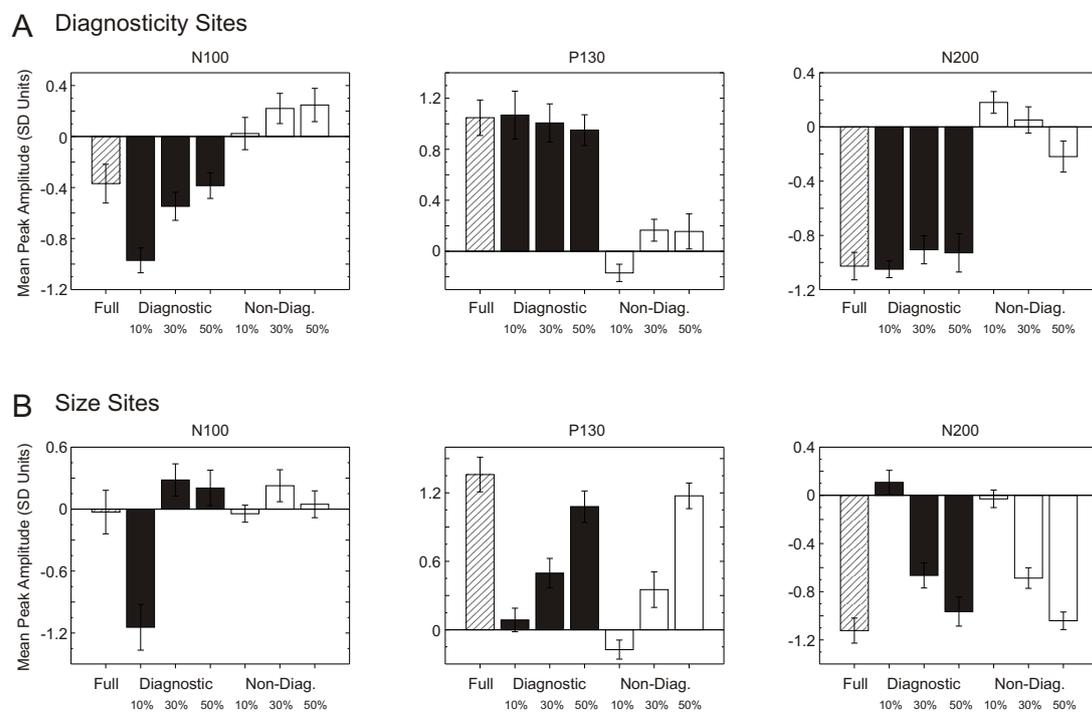


Figure 3.41: Amplitudes for the three VEP components. VEP amplitudes were computed for selected cases from monkey G00. Between 14 and 16 cases contributed to the averages. For a legend, see Figure 3.40.

Table 3.3: Number of diagnosticity and size cases in the medial and lateral parts of the recording regions for the two monkeys.

		G00		B98	
		Diagnosis- tivity	Size	Diagnosis- tivity	Size
N100	M	5	13	2	1
	L	9	1	6	7
P130	M	4	11	3	1
	L	10	3	5	7
N200	M	0	13	5	1
	L	14	1	3	7

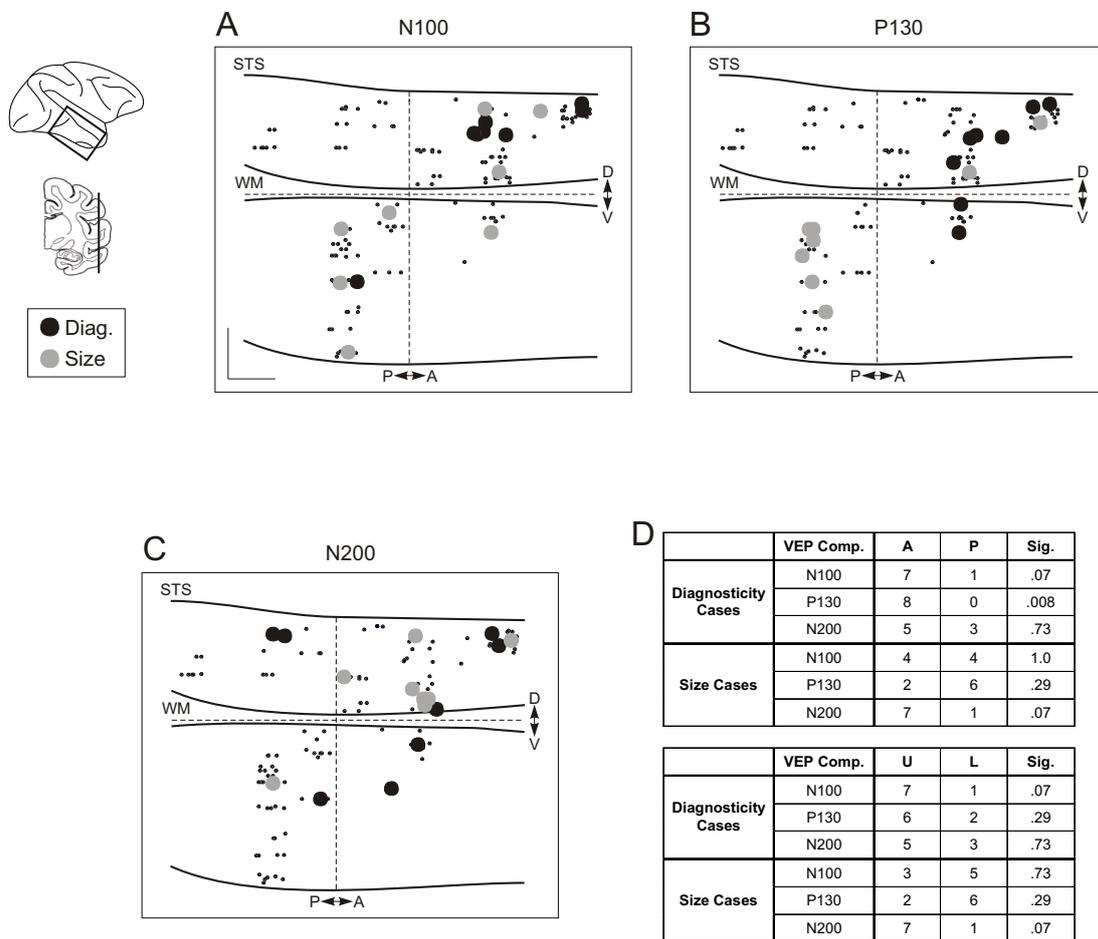


Figure 3.42: Distribution of responsive LFP cases. A-C, Topographies for the three time intervals. The plots show the location of diagnosticity and size cases for monkey B98 on sagittal views of the recording region. As in Figure 3.32, the view was generated by cutting a brain slice along the red line shown in the inset. Large dots correspond to cases classified as diagnosticity or size cases. Small dots show locations of the unclassified cases. The AP position of each case was randomly jittered by a small amount to allow a better separation of individual cases. Dashed lines indicate how the recording region was divided into subregions. Scale bars are given in the lower left corner of A. The two scale bars have a length of 1 mm. D, Number of cases in each of the subregions. p-values give the results of a binomial test between two subregions. Labeling of the subregions is shown in A-C, abbreviations are listed in Figure 3.32.

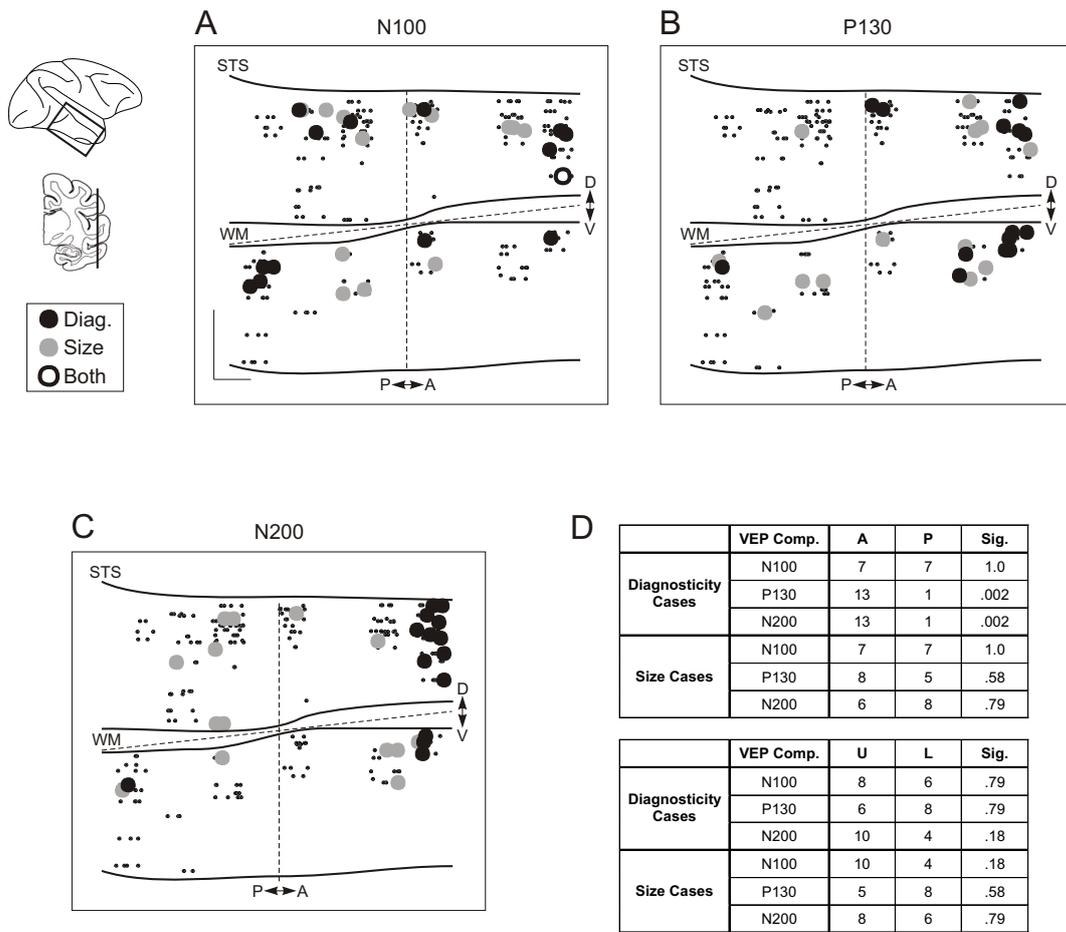


Figure 3.43: Distribution of diagnosticity and size cases for monkey G00. For a legend, see Figure 3.42.

of cases of the two categories was then counted in each of the subregions. These numbers are given in the tables in Figures 3.42D and 3.43D. A binomial test was used to assess whether the occurrence of recording cases with a particular property was significantly biased towards certain subregions. The results of this test are also listed in the tables in Figures 3.42D and 3.43D.

For the diagnosticity cases, a clear clustering in more anterior recording locations emerged. The results were most consistent between monkeys for the P130. For this component, the diagnosticity cases were clearly clustered in the anterior recording locations, with significant differences between the number of cases in the anterior and posterior regions in both monkeys. Results were somewhat more complicated for the other two VEP components: For the N100, diagnosticity cases clustered anteriorly for monkey B98, but were homogeneously distributed for monkey G00. In the case of the N200, the reverse was the case, with a clear clustering in monkey G00, and a more homogeneous distribution in monkey B98. However, despite the variability, there was no incident with more diagnosticity cases in the posterior than the anterior half. With respect to the effects of recording depth, there was a trend of diagnosticity sites to cluster in the lower bank of the STS; however, the comparisons did approach significance only in the case of the N100 component for monkey B98. Similarly, there was no consistent clustering of the size cases, which seemed to be equally present for all recording subregions.

The described results have been generated from the distribution of only the cases strongly influenced by the stimulus manipulations. To quantify trends present on the population level, the explained variances of all cases were plotted as a function of their AP position (see Figure 3.44). To test whether the results were consistent between monkeys, their data were combined. The resulting plots confirmed that the more anterior a recording site was located, the more the LFP amplitudes differed between diagnostic and non-diagnostic conditions. For all three time windows, a positive correlation emerged between AP position and the variance explained by the factor diagnosticity. This was the case for each monkey individually, as well as for the combined data. Correlations were significant when computing them across monkeys ($r \leq .008$ for the three time windows), as well as in most cases for each monkey individually (G00: N100: $p = .97$, P130: $p < .000$, N200: $p < .000$; B98: N100: $p = .03$, P130: $p = .08$, N200: $p = .001$). In contrast, the variance explained by changes in the visible stimulus size did not correlate with AP position. It also differed consistently between monkeys, with lower values in monkey B98 than in monkey G00. Correlation coefficients were in most cases not significantly different from zero as long as the monkeys were tested separately (G00: $p \geq .27$ for all three time windows; B98: N100: $p = .72$, P130: $p = .07$, N200: $p = .28$). When computing the

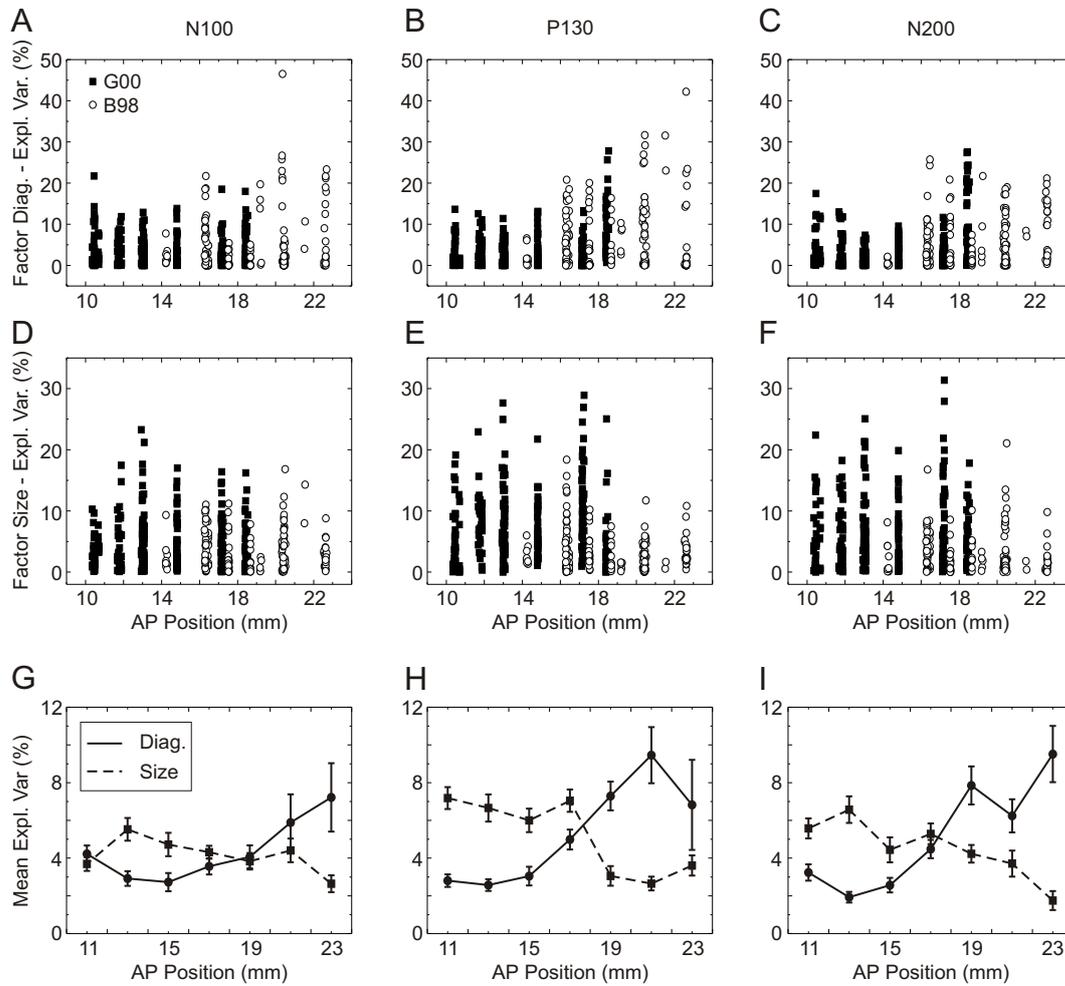


Figure 3.44: *Dependency of the explained variance on AP position. The results from both monkeys have been collapsed. A-C, Variances explained by changes in diagnosticity for the N100, P130, N200. D-F, Explained variances for the factor visible stimulus size at the three time windows. Plots summarize all visually responsive cases. The plots in A-F show the data of all cases. To confirm the observed trends, the recording region was divided into 2 mm wide bins, and the mean variances within these bins were computed. Mean variances are shown in G-I.*

correlation between the explained variance for the visible stimulus size and the AP position across monkeys, the lower values in monkey B98 in combination with their overall more anterior locations generated significant negative correlations for the P130 and the N200 (N100: $p=.29$, P130 and N200: $p<.000$).

In conclusion, the LFP behavior changed in a characteristic way when moving from posterior to anterior recording locations. Along this axis, the influence of diagnosticity on the LFP systematically increased. No such trend was obvious for the influences of the visible stimulus size. The LFP behavior is therefore in contrast to the single unit behavior, for which no dependency on the recording location was observed.

3.2.6 Relationship between LFP and SUA

Two neural signals were simultaneously recorded at each electrode – the spiking activity of single neurons, and the local field potential. While the LFP is a mass signal reflecting the synaptic events in a larger neural population surrounding the electrode tip, the spiking activity captures the output of a few, isolated neurons very close to the electrode. Since the signals were recorded simultaneously, they could directly be compared, allowing to assess how much the behavior of the LFP at a particular site correlated with the behavior of the single neurons recorded at the same site.

When analyzing the spiking activity, the firing rates in a time interval from 100 to 400 ms after stimulus onset have been used. In contrast, three time intervals centered around prominent VEP peaks were discussed for the LFP data. Before further discussing relationships between LFP and single unit activity, the timing of the VEP peaks was compared to the timing of the neural responses. Pooling across images, the onset latencies of the visually responsive neurons were computed separately for all conditions. Latencies were defined as the first 10 ms bin in which the number of spikes significantly exceeded the baseline level, followed by at least one equally significant bin. Figure 3.45 plots the determined distribution of latencies. The results for diagnostic and non-diagnostic conditions with the same visible stimulus size were pooled, to allow a direct comparison to the peak latencies determined for the VEP. VEP peak latencies are also plotted in Figure 3.45. As the graph shows, more than 25% of the neurons were already active at the time of the N100. The occurrence of the N200 fell at or after the onset latency of 75% of the neurons. The most interesting relationship was observed between the latency of the P130 and the neural latencies: Figure 3.45 shows that the median response latency corresponded

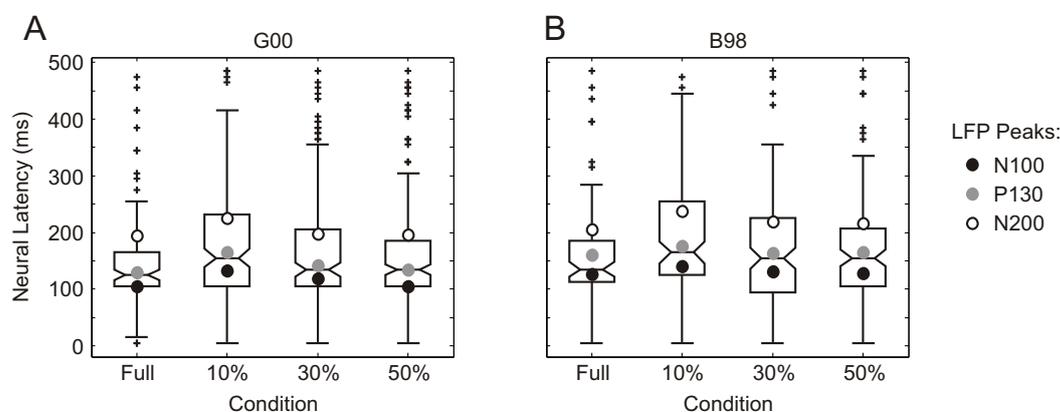


Figure 3.45: Comparison between neural latencies and the latencies of the three VEP components. Box plots show the distribution of latencies observed in the population of single units. The upper and lower end of each box are placed at the upper and lower quartile of the distribution, respectively. The middle bar indicates the median of the distribution. Overlaid on each box plot are the latencies of the N100, P130, and N200. A, Distributions for monkey G00. B, Results for monkey B98.

well to the peak of the P130. As the P130 latency, the median response latency was prolonged when only small stimulus portions were visible. It also showed the same shift of about 20 ms across monkeys.

To compare the functional properties of LFP sites and single units, two analyses were carried out. In each of these analyses, the characteristics of a LFP site were compared against the characteristics of the single units recorded at the same site, i.e. at the same electrode. If multiple units could be isolated at an electrode, the LFP results were tested against each of these units. In the first analysis, the influence of the different experimental conditions on the LFP and single units were compared. The second analysis more generally compared the stimulus selectivity of the two neural signals.

In Sections 3.2.3 and 3.2.4, the explained variance was used to quantify how much either firing rates or LFP amplitudes were influenced when the diagnosticity of a stimulus was changed, or the visible stimulus size was varied. These measures could be directly compared between LFP and single units. As described in Section 3.2.3, the single unit behavior was found to be very homogeneous throughout TE. In contrast, more anterior LFP sites were more strongly influenced by diagnosticity than more posterior sites. These different topographies already suggest a low correlation between the behavior of neurons and the LFP. This hypothesis was confirmed by plotting the explained variance data for the single units against the data for the

LFP amplitudes. Data were pooled over images. As before, responses of a LFP site were only included if at least one condition of an image evoked significant responses at this site. However, all single neurons were considered, irrespective of whether they responded significantly to a visible stimulus or not.

Figures 3.46 and 3.47 show the results for monkey B98 and G00. These plots also show the correlation coefficients between LFP and single unit data, and their significance. Irrespective of whether all or only the visually responsive single neurons were considered, there were no apparent correlations between the LFP and the single unit properties. The only exception was the N100 time period in monkey B98, in which LFP and single units recorded at the same electrode were influenced with related strength by changes in the factor diagnosticity.

Comparing the distribution of the explained variances for the visually responsive single unit cases against the distribution of the corresponding LFP values, higher values were observed for the single units. Paired t-tests between the variances explained by diagnosticity for single units and LFPs were significant at all three time windows for monkey G00, with a higher mean explained variance for single units than for the LFP ($t(67) \leq -2.95$, $p \leq .004$ for the three time windows). For monkey B98, the same trend was observed in the mean, but the difference did not reach significance at any of the time windows ($t(33) \geq -1.12$, $p \geq .27$ for the three comparisons). With respect to the factor size, the mean explained variance was again higher for the single units than for the corresponding LFPs. This difference was significant for the N100 time period in monkey G00 ($t(67) = -3.11$, $p = .003$), as well as all time periods in monkey B98 ($t(33) \leq -2.10$, $p \leq .04$).

Occlusion differently affected LFP and single unit responses recorded at the same electrode. In a second analysis, the selectivity of LFP and single units was therefore compared. This provides a handle on more general response properties of the two signals. To quantify the selectivity, only the responses to the unmasked images were taken into account. For the LFP sites, mean amplitudes during the N100, P130, and N200 were considered as response. The firing rate was taken as the single unit response. Selectivity was then measured as the variance in the response strength that could be attributed to differently strong responses to different images. The explained variance for the factor image was thus computed.

As before, the explained variance values for the single unit were plotted against the ones for the LFP amplitudes measured at the same electrode (see Figure 3.48). Only LFP cases were considered in which at least one unoccluded image evoked significant responses. All single units were again considered. Individual neurons were taken to be visually responsive if at least one unoccluded image evoked significant

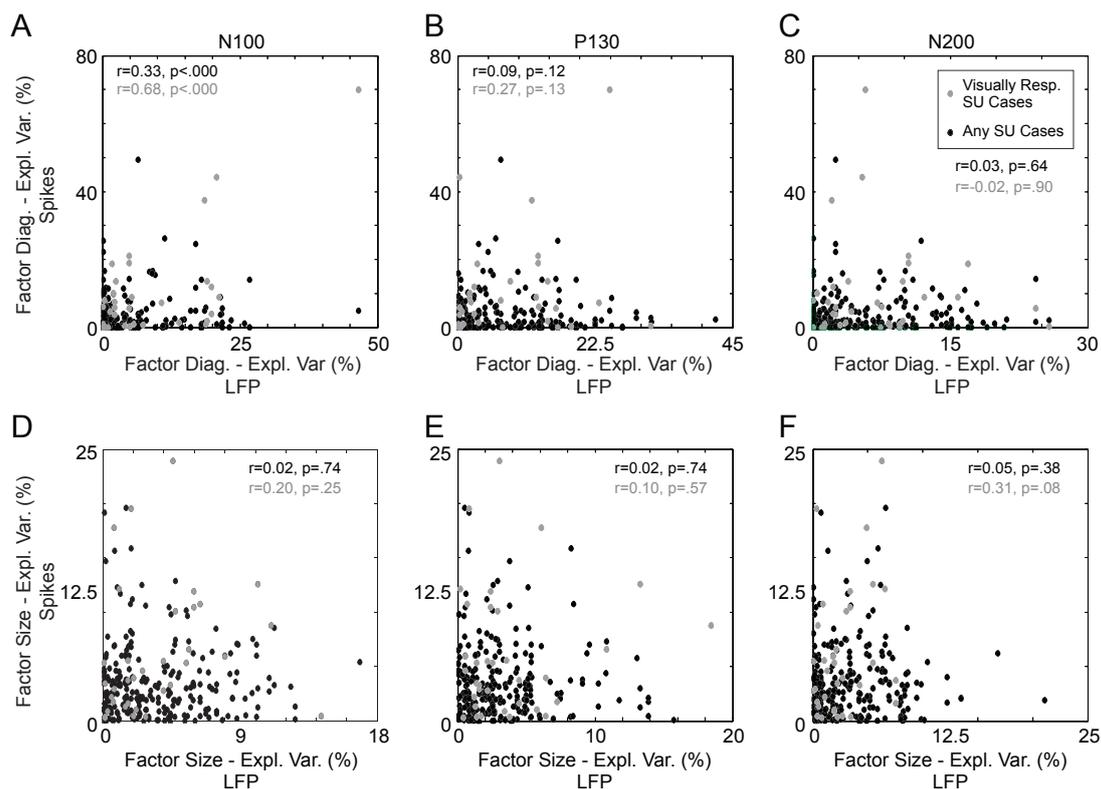


Figure 3.46: Comparison between single unit and LFP properties for signals recorded at the same electrode (monkey B98). The plots compare the explained variance data for the single units against the data for the LFP sites. The upper row (A-C) contains the variance explained by the diagnosticity, the lower row (D-F) the variances explained by the visible stimulus size. Each plot corresponds to one VEP component. LFP cases were restricted to visually responsive cases. Black dots indicate that the single unit case was not visually responsive; gray dots are visually responsive single unit cases. Correlation coefficients computed for the LFP and single unit data are given in each plot, together with their significance. Black values give the correlation coefficients for the complete data set, gray values the correlation coefficients considering only visually responsive single units.

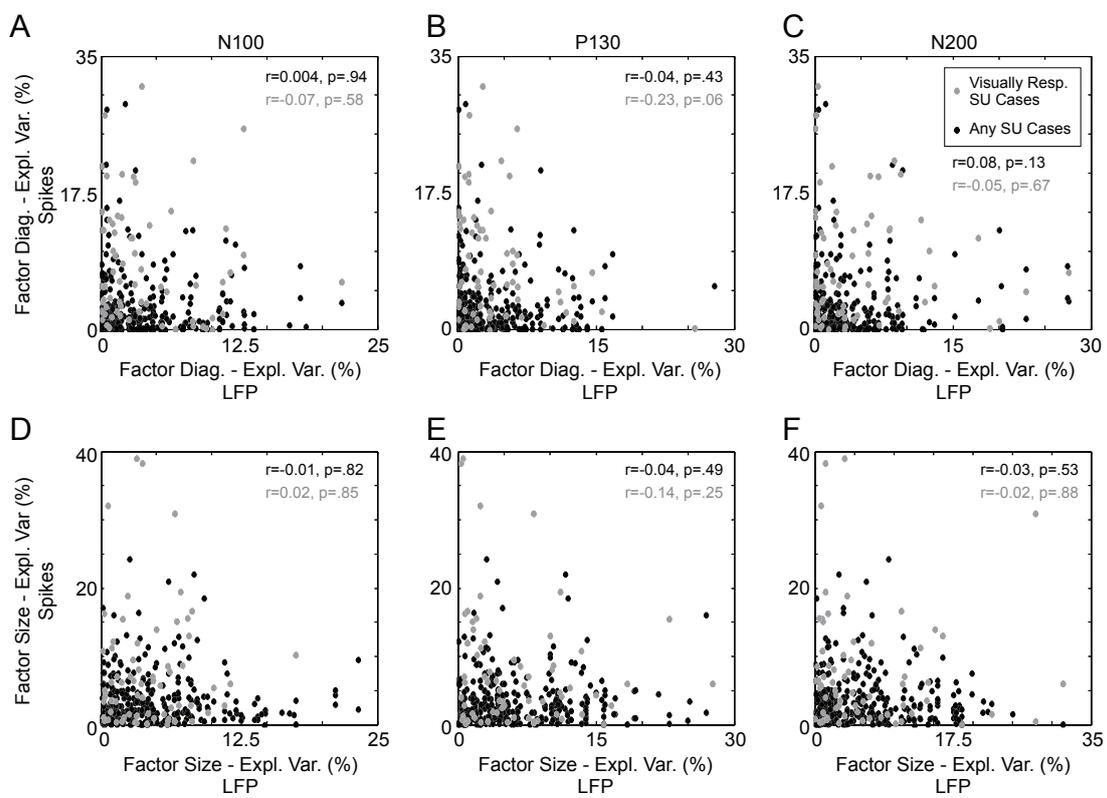


Figure 3.47: Dependency between single unit and LFP properties for monkey G00. For a legend, see Figure 3.46.

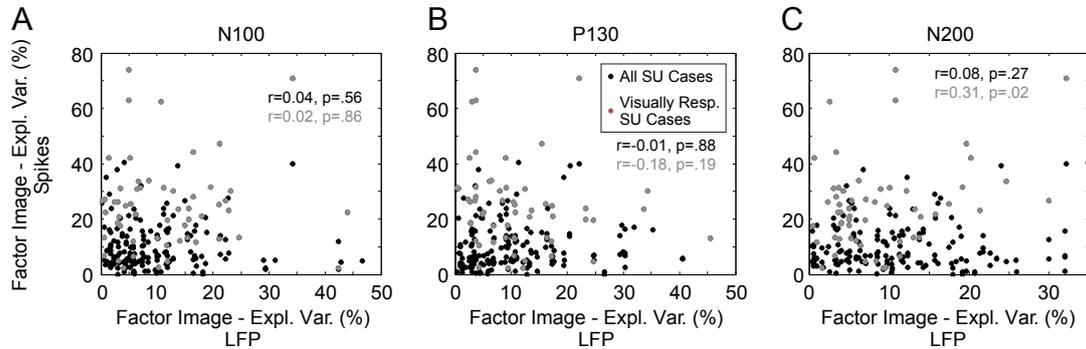


Figure 3.48: *Selectivity of single units versus the selectivity of LFP sites. Selectivity was quantified as the variance in response rate explained because of the presentation of different images. Only the unmasked images were considered. LFP sites were restricted to the ones showing visual responses. A-C, N100, P130, and N200. Black dots show the behavior of these LFP sites compared to the behavior of the single units recorded at the same location, with the single units being not visually responsive. Gray dots show the comparisons with visually responsive single units. Results from both monkeys were pooled. The inset text gives the correlation coefficients between the LFP and spike data (black: complete data, gray: significant single units only).*

responses. Effects were similar in both monkeys, and their data were therefore pooled, so that the analysis was based on 212 LFP cases. Again, no correlation became evident for the N100 and P130. In the case of the N200, the selectivity of the responsive single units was significantly correlated with the LFP selectivity. Comparing the range of explained values observed for the LFP and the single units, an even stronger tendency for higher values for the single units became obvious. Paired t-tests between the explained variances for single units and LFPs showed these differences to be significant at all time windows ($t(211) \leq -3.33$, $p \leq .001$ for the three tests). The same was the case when only including the visually responsive single unit cases ($t(55) \leq -5.33$, $p < .000$).

For spikes as well as for the LFP, explained variances were computed on the raw data. Firing rates were not corrected for baseline activity, and LFP recordings were not corrected for stimulus uncorrelated fluctuations. These random influences will reduce the explained variances. If one assumes that the contributions of spontaneous activity were generally lower for the spikes than the LFP, this might well explain why explained variance values were generally lower for the latter signal. In the final analysis, the selectivity of LFP sites and single units was therefore quantified based on the VEP amplitudes and the net firing rates. N100, P130, and N200 were analyzed separately. The analysis was carried out identically for single units and

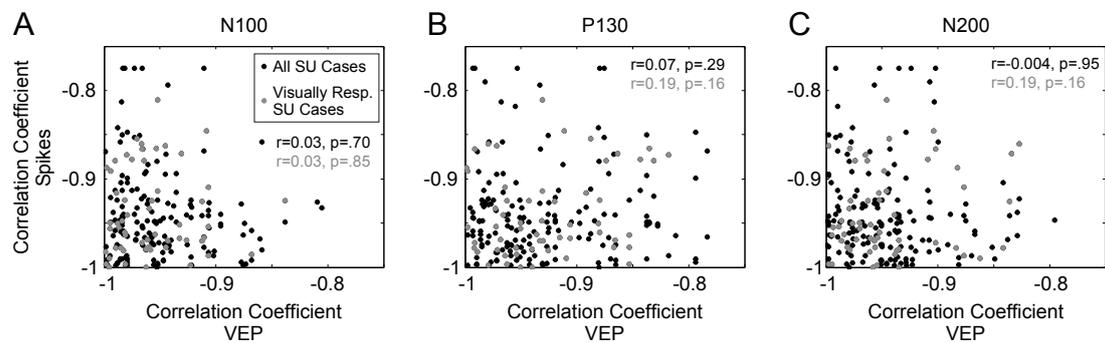


Figure 3.49: *Selectivity of single units versus the selectivity of LFP sites. Selectivity was here quantified by how monotonic a tuning curve decreased with stimulus rank. This monotonic decrease is measured by the correlation coefficient between the response of a neuron or the VEP amplitude of a LFP site, and the stimulus rank order. A-C, N100, P130, and N200. As before, black dots show the behavior of these LFP sites compared to the behavior of the single units recorded at the same location, with the single units being not visually responsive. Gray dots show the comparisons with visually responsive single units. Results from both monkeys were pooled. The inset text gives the correlation coefficients between the LFP and spike data (black: complete data, gray: significant single units only).*

LFP sites. The four images were rank ordered according to the response that they evoked when presented without an occluder. The best image was assigned to rank 1, the worst to rank 4. The responses were then correlated against the rank. This tests whether responses linearly decrease from best to worst image. The regression coefficient gives a measure for the monotony of a tuning curve. It underestimates selectivity, because very selective neurons or sites may only respond to one of the images, and therefore show no monotonic tuning curve.

Figure 3.49 plots the correlation coefficients of LFP sites against the ones for single units. Confirming the explained variance results, no correlation became evident between LFP and single unit behavior. However, when selectivity was characterized by the monotony of the tuning curves, LFP sites were as selective as the single units. 108 single units had significant correlation coefficients ($p < .05$); for the three VEP components, correlation coefficients were significant at 144, 101, and 122 LFP sites, respectively.

To compute a tuning curve, stimuli had to be rank ordered according to the response which they evoked. This rank order could be used to assess whether the image that evoked the highest responses from a neuron similarly generated the highest VEP amplitudes. To this end, each neuron's best image was selected. The rank of this image was then determined for the corresponding LFP site based on the

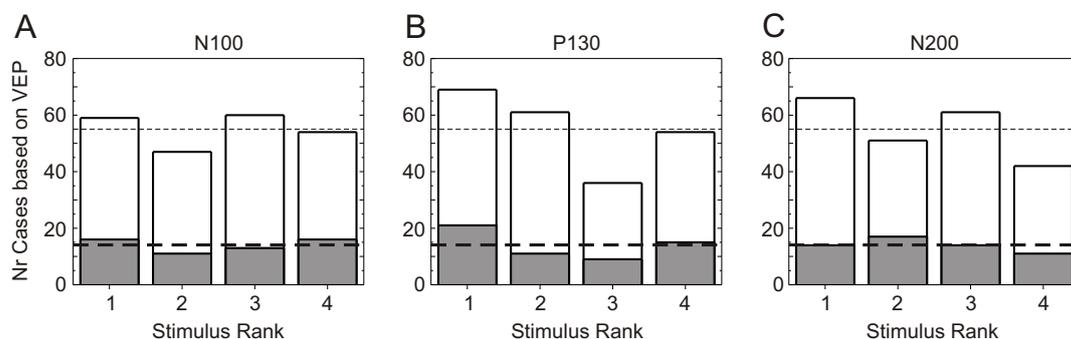


Figure 3.50: *Similarity of the stimulus preferences between LFP and single units. For each neuron, the best image was determined. It was then tested which rank this image occupied in the tuning curve of the LFP site, based on the VEP amplitudes. A-C, VEP components. Open bars sum all cases, gray bars the visually responsive neurons only. The thin dashed line indicates chance level for all cases, the thick dashed line the chance level for the visually responsive ones.*

VEP amplitudes. A rank of 1 implies that the neuron’s preferred image was also the best stimulus for the LFP sites. Larger ranks indicate that the LFP site did not preferentially respond to the neuron’s best stimulus. The histograms in Figure 3.50 show the distribution of ranks at the three VEP components. These plots clearly demonstrate that all rank positions were about equally likely. There was no indication of a similar image preference for neurons and the corresponding LFP sites. To confirm the significance of the observation, the number of rank 1 cases was tested against the chance level expected for an equal distribution across all ranks. Irrespective of whether the complete data or only the visually responsive neurons were considered, there was no significant deviation from chance level (χ^2 -test, $\chi^2 \leq 3.56$, $p \geq .06$). In conclusion, the selectivity of single units and LFP sites was very different.

All analyses performed in this section highlight that the properties of the LFP recorded at an electrode are not correlated with the properties of the single neurons recorded at the same electrode. This was the case for the effects of occlusion, as well as the general stimulus selectivity.

Chapter 4

Discussion

The aim of this Ph.D. thesis was to characterize how occlusion influences the representation of learned visual form by neural populations in the temporal lobe. In an initial, purely behavioral study, the influence of different spatial placements of an occluder on the recognition of monkeys was tested. The results of the behavioral part indicate that monkeys rely more on some parts of images than on others. These results were used to study the responses of neurons in area TE. Here, changes in firing rate and LFP responses with different occluder placements were analyzed. The results show how the diagnosticity of an image region influences neural responses measured in area TE. The discussion will first concentrate on the behavioral results, followed by the conclusions drawn from spike and LFP data.

4.1 Behavioral results

4.1.1 Effects of occlusion on the performance of monkeys

In most cases, monkeys were able to identify partially occluded images. However, despite their overall good performance, occlusion introduced recognition errors. The effects of a mask were dependent on which image parts were occluded: Whereas the occlusion of certain image regions consistently impaired performance, the occlusion of others had little or no effect. This is the first report of such effects for Rhesus monkeys. Previous studies have tested humans and pigeons with the Bubbles paradigm. They have similarly shown that the placement of the occluder determines whether an image could be identified (Gosselin & Schyns, 2001; Gibson et al., in

press). It therefore seems that partial occlusion affects the perception of different species in a similar way.

The observed systematic interactions of occluder and recognition performance imply that the monkeys preferentially used information from specific image regions to perform their task. We could thus successfully determine the diagnosticity of each region in the natural scenes. The current study is the first in which this was possible. Because of the complexity of natural scenes, diagnostic regions in natural scenes could so far only be indirectly determined. They were either deduced from the errors in classification tasks (D'Amato & van Sant, 1988; Macé et al., 2005), from the groups into which large numbers of stimuli were classified (Sands et al., 1982), or from the behavior towards manipulated images (Vogels, 1999). These methods relied on the intuition of the experimenter about which features could be diagnostic. Furthermore, none of the studies could determine with certainty all diagnostic regions. Here, diagnostic regions were determined with the actual stimuli. The paradigm did not require any prior expectations about stimulus diagnosticity, and exhaustively sampled all image regions. The current study therefore provides a much more complete picture of the monkeys' behavior than the previous ones.

The monkeys differed markedly in which image regions they found diagnostic for the task. Both monkeys only partially agreed with human observers about the diagnosticity of different image regions. Thus for the tested task, the behavior of human observers could not be used to predict the monkeys behavior. It seems that in human observers, the bias to use object parts to identify an image is much stronger than in monkeys. It can be speculated that these differences are generated because the observers use different levels of abstraction to encode the stimuli.

When recording the eye movements evoked by an image, the distribution of fixations has been taken to indicate which image regions an observer finds informative. It has been shown that the fixation density for an image changes depending on the information that an observer needs to extract from the image. This result was confirmed for monkeys. Fixation densities changed when the monkey was performing a task on an image, instead of freely inspecting it. Nonetheless, neither the fixation locations during free viewing, nor the ones during the task necessarily agreed with the Bubbles results.

A number of reasons may explain the discrepancy: First, it is known that at least the locations of the initial fixations depend on low-level image properties (see Section 3.1.4). However, it is unclear how the influence of low-level properties on fixation placement develops over time. Furthermore, the monkeys were not forced to look at the image the whole time. As a consequence, they usually studied the image for

some time, but eventually looked away from the image. These periods of looking elsewhere were often followed by another period of fixations falling onto the image. Again, it has not been tested how the physical image characteristics influence these “rebound” fixations. In conclusion, it is almost impossible to quantify how much the fixations should be influenced by the low-level image properties. If they strongly determine the location of fixations under the tested settings, this could explain why diagnostic image regions did not necessarily receive most of the fixations.

It is also possible that the placement of fixations on an image is generally very stereotyped. Low-level image properties could be one source determining the stereotyped behavior, but other sources like image content are conceivable as well. Even though the task seems to exert some influence on the placement of fixations, this effect may not be strong enough to completely overcome the stereotyped behavior. A closer inspection of Figure 3.13, which sparked the study of fixations, supports this hypothesis. Although the distribution of fixations varied greatly depending on the instruction given to an observer, there was nonetheless no condition in which the portrayed heads received no fixations. When for the tested monkey the fixation densities were compared between a free-viewing and a task condition, it was found that the same image regions received most of the fixations. Yet, there was a trend for the diagnostic image regions to receive more fixations in the task condition. Therefore, it could be hypothesized that the image regions attracting similar numbers of fixations in both paradigms reflect the stereotyped component of the fixation placement. The additional image regions receiving fixations only during the task could then result from the small task influence. Precisely these image regions showed more consistency between the eye movement behavior and the Bubbles results.

Finally, one could conclude that monkeys identify partially occluded stimuli differently than fully visible stimuli. Since only the eye movements evoked by unoccluded images have been quantified, this could explain why image regions with high fixation densities did not predict the diagnostic image regions well. It seems very unlikely that the monkeys identify occluded and unoccluded images based on different image regions (see also the discussion below). This hypothesis could possibly be verified based on the scan patterns evoked by partially occluded stimuli. However, fixations usually do not fall onto homogeneous parts of an image. The occluded image regions do not contain structure, and it is thus unlikely that they would attract fixations. The fixation density on a partially occluded image therefore has to be highest for the image regions that are visible through the occluder. However, this effect does not necessarily indicate a different treatment of the partially occluded stimuli. It can simply be a product of which image regions contain visible structure that can attract fixations.

The diagnostic image regions for the Bubbles task could neither be predicted from the behavior of human observers, nor from the fixation densities. Predictions based on low-level image properties were similarly unsuccessful. For monkeys, no differences could be found in the average low-level properties of diagnostic and non-diagnostic regions. Also, even in the best case, only about 40% of a diagnostic region could be explained by the occurrence of extreme values of some low-level characteristics. About the same portion could be explained for the non-diagnostic image regions.

Taken together, no parameters have emerged that can predict the behavior of the monkeys. Most importantly, even though monkeys are often assumed to behave similar to human observers in visual tasks, the current results suggest that inferring monkey behavior from human behavior can be very misleading. This emphasizes the need to directly determine the strategies of monkeys in a particular task. The observed variability between monkeys strongly supports this claim. While monkey G00 used mostly object parts to identify a scene, monkey B98 relied on background patches to identify the same scene. Based on more standard measures of an observer's performance, this difference in strategies would not have become apparent. As an example, both monkeys performed with a comparable, very low error rate when they had to discriminate among unoccluded images.

The current study is not the first to suggest that good performance in a task does not indicate that monkey observers are behaving similar to human observers. Studies on stimulus categorization both for monkeys and pigeons draw similar conclusions. D'Amato & van Sant (1988) trained monkeys to categorize natural scenes based on whether they contained a person. The tested monkeys seemingly learned the concept very quickly and could apply it to novel scenes. After a series of experiments, the authors analyzed the pictures that were misclassified by the monkeys. Based on the pictures that were erroneously classified to contain a person, it became evident that in many cases the monkeys were actually using the presence of a red patch to identify natural scenes that contained a person. This strategy is in strong contrast to the behavior of a human observer performing the same classification.

Prior to the study by D'Amato & van Sant, Greene (1983) similarly demonstrated for pigeons that little about the birds' strategies could safely be concluded from their obviously good performance in a task. In this study, pigeons were trained in a number of complex categorization tasks, including a categorization task involving slides with or without a particular person. In all tasks, it became evident that the pigeons achieved a good performance because they memorized the large stimulus sets, not because they extracted the categorization rules. Interestingly, for the classification of the person slides, the results are reminiscent of the behavior of

monkey B98. In the initial training, person and non-person slides had distinctly different backgrounds. When during testing the backgrounds were swapped between the two categories, pigeons failed to respond to the person slides as before, and instead responded to the non-person slides. Presumably, pigeons had based their decisions on the background instead of the person shown in the scenes.

4.1.2 Methodological issues on Bubbles

The present results were obtained using the Bubbles paradigm, by determining the image regions where occlusion had a systematic effect on performance. For completeness, this section discusses potential problems with the Bubbles paradigm, and their consequences on the experimental results. This is also necessary because Murray & Gold (2004) recently questioned the validity of the Bubbles approach. They criticized two aspects of the paradigm. First, they suggested that reverse correlation is more appropriate to determine diagnostic regions than Bubbles. Second, they claim that Bubbles distorts the diagnosticity of image regions and therefore yields incorrect results.

The first issue is irrelevant for the current project. Reverse correlation in principle allows the same type of analysis as Bubbles. In a typical reverse correlation experiment, random noise drawn from a Gaussian distribution is added to each image pixel (see for example Gold et al., 2000). Similar to Bubbles, the observers continue to perform their task on these noisy images. After the experiment, image regions are identified in which the addition of noise interferes with the observers' behavior. In reverse correlation, each pixel is treated independently. In a given trial, very different noise values can thus be added to neighboring pixels. As a consequence, the appearance of the complete stimulus resembles the one of an image received on a TV channel that is out-of-tune. Although certain stimulus information is deleted in the noisy image, the stimulus does not generate the impression of a partially occluded image. Since the goal of the current project was to study the influences of partial occlusion on behavior, reverse correlation was not suited for the project.

The second issue raised by Murray & Gold however is more relevant. They argued that because only small fragments of a stimulus are visible on any Bubbles trial, observers will treat these stimuli differently than the unoccluded stimuli. If observers change their strategies for the occluded stimuli, the Bubbles results will not truthfully capture the diagnostic regions in the original stimuli. As an example, since natural scenes contain redundant features, it may be that observers normally use only few of these. Presenting the natural scenes behind occluders may then force

the observers to use different features on different trials, depending on how the mask is placed. As a consequence, the diagnostic regions obtained with Bubbles would include all possibly informative image regions, not only the fewer image regions usually used. Another potential influence of partial occlusion has been discussed in Section 3.1.3. Under normal circumstances, at least human observers identify each image by its foreground object. When images are partially occluded, the lack of information may make this high-level strategy less successful. In turn, the observers may begin to use more low-level strategies to perform the task. Image regions with prominent low-level features do not have to coincide with image regions containing object information; the diagnostic regions determined with Bubbles would therefore be a misrepresentation of the diagnostic regions in the unoccluded image.

Based on our experimental data, it seems unlikely that Bubbles strongly interfered with the subject's normal strategies. All subjects, both humans and monkeys, were immediately able to perform the task on the partially occluded images. Strategies thus must have been transferred from the unoccluded images. In addition, the diagnostic regions remained very stable. As has been shown for monkey B98 in Section 3.1.2, the regions determined as diagnostic after prolonged testing were already diagnostic during the first session. Thus, diagnostic regions were not a product of repeated exposure to masked stimuli. This also means that subjects do not adapt their behavior to the partially occluded stimuli. The two observations can indicate that subjects use the same strategies for occluded and unoccluded images. They are, however, also compatible with an immediate change in strategy as soon as occluded images are introduced. This hypothesis cannot be fully excluded. However, since masks were random on every trial, it seems unlikely that subjects could adapt so quickly to the new situation. Finally, a pronounced influence of more low-level image features during the Bubbles testing can be refuted as well. For none of the observers could the prominent presence of low-level factors explain more than 40% of the diagnostic regions. Image regions did thus not become diagnostic because subjects preferentially concentrated on these features when tested with partially occluded stimuli. In conclusion, our experimental data suggests that the same regions are diagnostic for occluded and unoccluded images.

Most importantly, even if Bubbles would change the strategies of an observer, this would only affect the interpretations of the behavioral data. It would have no consequences for the neurophysiological part of this project. The aim of this thesis was to characterize the effect of occluder placement on the responses of area TE neurons. For the neurophysiological data, the main comparisons were performed between different occluded versions of an image. Responses to unoccluded images served only as a benchmark, to quantify general influences of occlusion. To this end, the goal of

the behavioral studies was simply to identify image regions where occlusion affected the monkey's performance. These tests allowed us to generate interesting conditions for the single cell recordings. The Bubbles paradigm was used since it performs a rigorous sampling of the whole image space without imposing any biases for specific image regions. As confirmed in a separate experiment, the obtained results indeed could be used to construct occluded image versions that affected the monkeys' performance. With respect to the physiology study, the Bubbles paradigm thus clearly served its purpose.

The discussion so far has dealt with specific issues regarding Bubbles. It was initiated by the concern that the used paradigm influences the strategies of observers, and can thus not capture the diagnostic regions of the unoccluded images. These effects may not be present, and are furthermore not important for the project. They are however one aspect of a general problem regarding diagnostic regions: Diagnostic regions are not fixed, but depend on the requirements of a task. This general fact, which is not a product of the Bubbles paradigm, has relevant implications for the current project. First, the determined scene regions are only diagnostic for the tested task. The validity of this argument can be inferred from the studies of Gosselin & Schyns (2001) and Gibson et al. (in press). In both experiments, the diagnostic regions changed when subjects performed different tasks. It is therefore likely that if our identification task was replaced by a different task, diagnosticity would be distributed differently across the scenes. Second, the diagnostic regions depend on the stimulus set. Here, each stimulus set consisted of three images. Adding further images, or pairing the tested images with different scenes, would probably change the diagnostic regions. As a thought experiment, assume that scene N1 would be paired with two scenes showing the same bird in the same pose, but in front of different backgrounds. Necessarily, the background would have to become diagnostic instead of the bird. Because of these two reasons, the determined diagnostic regions cannot be taken as fixed entities in the natural scenes, but they should be interpreted only in the context of the task and the stimulus set. These regions reflect the learning of both task contingencies and stimulus properties.

These caveats should be kept in mind when interpreting any neural effects as well. It cannot be assumed that images will be split in the same diagnostic and non-diagnostic regions under all circumstances. As a consequence, it seems unlikely that the responses of TE neurons would show such a rigid distinction between different image regions. This implies that, in principle, the responses of TE neurons to diagnostic and non-diagnostic conditions have to be recorded while the monkey performs the appropriate task on the correct stimulus set. The data reported here were collected while the monkeys performed a fixation task; in addition, the stimuli

from different sets were mixed. However, because of the large amount of trials that needed to be collected for the Bubbles analysis, the monkeys had been extensively exposed to the stimuli prior to the recordings. During the behavioral testing, each stimulus appeared only in one context. As a consequence, the monkeys developed a very stable preference for certain image regions. Indeed, the re-test of monkey B98 on the first image set showed that even training on a second image set did not influence the diagnostic regions. As discussed in Section 1.2.6, the selectivity of TE neurons is shaped by learning. Because of the repeated exposure to the images and the presumably stable diagnostic regions within these images, it is very likely that TE neurons developed a matching selectivity. This explains why the responses of TE neurons showed a pronounced difference between diagnostic and non-diagnostic conditions even in the fixation task with stimuli from different sets.

A final methodological issue concerns the selection of parameters for the Bubbles paradigm. With respect to the occluders, two parameters had to be chosen. These were the number of bubbles in a mask, and the size of the bubbles. The number of bubbles varied within a broad range for monkey and human observers. Its influence on the Bubbles results therefore has to be negligible. The influence of the bubble size was directly tested for human observers. The results show diagnostic regions to be largely unaffected by a change in bubble size. For the monkey observers, the influence of the bubble size was not tested; however, it is unlikely that diagnostic regions would change drastically with different bubble sizes. When the Bubbles results were verified, the smallest masks revealed half of an image. In these stimuli, the visible image parts were largely continuous, a mask configuration that paralleled Bubbles masks with very large bubbles. Still, diagnostic and non-diagnostic conditions led to markedly different performances. Also, larger number of bubbles were used for monkeys than for humans. For the human observers the limiting effects of a smaller bubble size could be overcome by placing more bubbles in an occluder. Thus, for the monkeys any influences of a too small bubble size would probably have been counterbalanced by the large number of bubbles.

4.2 Single unit responses

Neural responses in area TE were recorded while the monkeys viewed different partially occluded versions of natural images. Six masked conditions were constructed for each image. Occluded conditions varied in two factors, diagnosticity and visible stimulus size. Both factors characterize different properties of the occluders. In the diagnostic conditions, image regions with behavioral relevance were visible; these regions were occluded in the non-diagnostic conditions. Diagnosticity thus refers to the spatial placement of the occluder. The visible stimulus size indicates the occluder's extent. Behaviorally, the two factors had very different effects: The diagnosticity of the visible image portion determined whether or not an image could be identified despite occlusion. In contrast, there were almost no influences of the visible stimulus size (see Figure 3.9). Decreasing how much of a stimulus was visible only slightly reduced the recognition rate for diagnostic stimuli. Also, with the exception of image set 1 for monkey B98, increasing the visible stimulus size did not improve the performance on the non-diagnostic conditions. In conclusion, the influence of occlusion on TE responses were studied with respect to two occluder parameters, one with behavioral consequences, the other one without.

The results of a previous study on occlusion (Kovács et al., 1995) showed how TE neurons were influenced by occlusion in general, irrespective of the spatial placement of the occluders. The previous study differed in a number of ways from the current one: Instead of natural scenes, geometric shapes were tested. Furthermore, occluders were either moving or static masks consisting of randomly placed texture elements. Before describing the effects of the occluder placement in more detail, it is thus of interest to test whether both studies agree on the general influences of occlusion on TE responses. For this comparison, the diagnostic and non-diagnostic conditions were pooled.

One effect of occlusion reported previously was a general reduction in firing rate. Kovács et al. (1995) quantified the effect of occlusion by a so called responsivity index (RI), which was computed by subtracting the net response to the preferred shape in an occlusion condition from the net response to the same shape without occlusion. The difference was then divided by the sum of the two responses. RI values near 0 show that responses were comparable with and without occlusion. Values near 1 are generated if a neuron is not responsive to the occluded shape at all. For a moving occluder covering 50% of a shape, Kovács et al. reported a median RI of 0.26, indicating that the response strength with partial occlusion was one-half the response strength with no occlusion. For a stationary occluder also covering 50% of a shape, a median RI of 0.17 was obtained, which was statistically

not significantly different from the RI with moving occluders. Increasing the density of the occluder, i.e. decreasing how much of a stimulus was visible, increased the RI. For a moving occluder covering 90% of a shape, the median RI reached 0.66.

Judging from Figure 3.23B, occlusion reduced the firing rates at least by a factor of two in the current study. The size of the response reduction reported by Kovács et al. thus seems similar to the one observed in our data. To compare the two experiments more directly, the RI was calculated for our data as well. The analysis was based on the usual pool of 220 visually responsive cases, and therefore not only included the preferred images, but any image that evoked a significant response. The responses to diagnostic and non-diagnostic conditions of the same size were averaged. Median responsivity indices of 0.37, 0.32, and 0.25 were obtained for the 10%, 30%, and 50% condition, respectively. Especially the RI for the 50% condition nicely agrees with the results of Kovács et al., despite the differences between the experiments. In the 10% case, the response reduction seems to have been less pronounced in our neuronal pool. However, with random occluders covering 90% of a shape, chances are low that the remaining 10% are the diagnostic shape parts. In the study by Kovács et al., most trials with the very dense occluder will therefore have been exposing non-diagnostic shape features. A fair comparison between the two experiments should thus probably be based only on the non-diagnostic condition of our study. Indeed, in our data the 10% diagnostic and non-diagnostic conditions evoked very different firing rates. Especially the responses in the non-diagnostic conditions were very low. Computing the RI for the non-diagnostic condition only, the median RI increased to 0.56, a value more consistent with the strong response reduction seen by Kovács et al.

Kovács et al. showed another general influence of occlusion on the neural responses. They reported that neural response latencies to occluded shapes were longer than the latencies for unoccluded shapes. For a static occluder with a coverage of 50%, the latencies differed by 20 to 25 ms. Considering Figure 3.45, occlusion also affected the latency in our study. To compare our data with the experiment of Kovács et al., latency differences were quantified between the unoccluded and the 50% condition, averaging latencies for the diagnostic and non-diagnostic conditions. Differences were computed for the 147 cases from both monkeys in which latencies could be determined for the unoccluded and occluded condition. A significant difference of 25.92 ms emerged, by which the responses to occluded images lagged the responses to unoccluded images (Wilcoxon signed rank test, $z=-3.01$, $p=.003$). Again, our data is in good agreement with the behavior of TE neurons reported previously.

Besides these general effects, occluders exerted different influences on TE responses depending on how they were placed. As described above, the mean firing rate was

reduced when images appeared behind occluders. However, the reduction in firing rate was less pronounced when diagnostic image regions remained visible despite the occlusion. The effect of occluder placement was strongest when only 10% of the image remained visible through the mask. In this condition, showing non-diagnostic instead of diagnostic image parts reduced the firing rate by about a factor of two. However, differences were also present for other visible stimulus sizes. To summarize, the responses of TE neurons followed the behavioral relevance of the image portions which remained visible after partial occlusion.

Differences in firing rates between diagnostic and non-diagnostic conditions could not be attributed to a number of possible confounds. Diagnostic and non-diagnostic regions covered different spatial locations. The influence of receptive field locations and inhomogeneities therefore needed to be considered. However, image material at a fixed spatial location could or could not evoke responses from a neuron simply depending on its diagnosticity. This clearly argues against an influence of the spatial position at which stimulus material was presented. Similarly, low-level image parameters could be excluded as a source for differences between diagnostic and non-diagnostic conditions. All occluded versions of an image had the same global luminance and contrast as the unoccluded image. In most cases, this also equalized luminance and contrast locally. In addition, the firing rates of only very few neurons significantly correlated with the low-level properties of an image, as shown by the analysis in Section 3.2.1. Finally, diagnostic regions – as had to be expected – usually contained some part of the object shown in the images. TE neurons are known to be responsive to object features. Differences between diagnostic and non-diagnostic conditions could therefore also be due to the presence or absence of an object, instead of differences in diagnosticity. Although it is very difficult to disentangle the influences of diagnosticity and objects, a limited analysis could be performed on the neurons from monkey B98 that responded to scene N2. To identify this scene, the monkey used information from a background region, and non-diagnostic conditions contained the scene’s object. In this case, no differences emerged between the diagnostic and non-diagnostic condition. The results therefore demonstrate that the presence of an object partially contributes to the differences between conditions. It is strong enough to counterbalance the influence of diagnosticity, so that responses to diagnostic and non-diagnostic conditions become similar. Yet, since the responses were not stronger in the non-diagnostic condition, the mere presence of an object cannot completely explain the differences in firing rates for diagnostic and non-diagnostic conditions.

Besides differences in mean firing rate, occluder placement also affected the selectivity of TE neurons. Nevertheless, neurons maintained more of their selectivity in

the diagnostic conditions than in the non-diagnostic conditions. The spike counts also transmitted more information about which image was presented in a diagnostic condition. The mutual information in the diagnostic conditions even reached the same values as in the unoccluded condition. Two conclusions can be drawn from these observations. First, based on partial occlusion with random masks, Kovács et al. (1995) concluded that TE neurons remained selective despite the occlusion. Our data show that this conclusion is limited: Selectivity is only preserved in the diagnostic conditions. Second, diagnostic regions are image regions that the monkeys used to discriminate among the different images. Therefore, it is interesting to see that neural responses are also better able to distinguish among diagnostic regions of different images, than among non-diagnostic regions. However, it should be kept in mind that selectivity was computed across four images selected from two image sets. While the monkeys had extensive experience in discriminating the images from the same set, they were never required to perform a comparison across image sets. It is understandable that selectivity developed for the members of an image set, but it is not as clear why the neurons should show selectivity across the members of different image sets.

In contrast to diagnosticity, the visible stimulus size only had a small influence on the neural responses. For the mean firing rates, enlarging the visible image portion only affected the non-diagnostic conditions. Here, firing rates increased when more of an image became visible. This effect is most likely due to the fact that with larger visible image regions, the non-diagnostic conditions may begin to include image regions with a somewhat higher diagnosticity. Neural selectivity was equally little influenced by the visible stimulus size. For the non-diagnostic conditions, increasing the visible stimulus size did not restore the selectivity found without occlusion. In the diagnostic conditions, selectivity increased with larger visible stimulus sizes. However, the major improvements were seen when increasing the visible stimulus size from 10 to 30%. Further improvements were small when moving from a visible stimulus size of 30% to one of 50%. Finally, the mutual information between spike count and image identity was unchanged by the visible stimulus size. With respect to neural selectivity, the selectivity in the diagnostic conditions became somewhat more pronounced when the largest amount of the image was exposed. As for the mean firing rates, this effect may be explained by the fact that larger non-diagnostic regions may begin to include more diagnostic image parts.

For an additional analysis, the influences of both stimulus manipulations were quantified by the size of firing rate modulations they could introduce. These modulations were captured in the variance explained by either changes in diagnosticity, or changes in visible stimulus size. Based on the explained variance data, it was possible to

study whether the neural behavior systematically changed with recording location. At least along the anterior-posterior direction, no influences of recording position could be identified for the single units: When plotting the AP recording positions of neurons strongly affected by one of the two factors, no clustering emerged in any of the monkeys. Since recording chambers were placed slightly differently in the two monkeys, the dependency of the explained variances on the AP recording position could be tested for a relatively large extent of TE. Again, no consistent pattern emerged for the two factors. In summary, the behavior of single neurons was relatively constant throughout the covered TE region. Most importantly, this also implies that different subregions in TE are similarly engaged in encoding stimulus diagnosticity.

In conclusion, the influences of occlusion on TE neurons parallel its behavioral consequences: Occluder placement determined whether or not a stimulus could be identified. These large behavioral effects were accompanied by large changes in firing rates and selectivity. In contrast, changing how much of an image could be seen had little influence on the recognition rate as long as the visible parts remained diagnostic. At the neural level, the addition of physical information introduced some changes, but these were of a different magnitude than the changes generated by diagnosticity.

Reducing an image to its diagnostic region diminished the firing rates somewhat; reduction to the non-diagnostic regions had a very pronounced effect on the firing rates, demonstrating that images can be split into regions that result in different TE responses. Since the responses were also reduced for the diagnostic conditions, all image regions seem to contribute to some degree to the response to the full image. However, diagnostic regions have more impact on the firing rate than non-diagnostic regions. This result in general could have been predicted from the experiments by Tanaka et al., described in Section 1.2.6. These studies reported that TE neurons responding to a complex object often also responded to specific parts of the object, but not to others. Yet, no unifying criterion for the critical features emerged from these studies. Our data show that the critical image features that drive TE neurons are the regions of behavioral relevance. This finding is consistent with a number of studies, all of which have shown that diagnosticity plays a large role for the responses of TE neurons. Along diagnostic dimensions, TE neurons express higher selectivity than along other dimensions (Sigala et al., 2002; Baker et al., 2002). Furthermore, during visual search, TE neurons remain silent to large parts of natural scenes, but begin to respond to the regions indicative of the appropriate response to the scene (Sheinberg & Logothetis, 2001).

The data also point to another property of TE neurons. The formation of diagnostic regions is a result of experience. Besides preferences resulting from general experience with visual input, there is no reason why an observer should, a priori, find one image region more diagnostic than another. Similarly, it is possible that TE neurons in naïve animals show some distinction between different image regions because of the general experience with visual input. Our data, however, show clear differences between the image regions that were diagnostic or non-diagnostic for the task. The fact that these differences were present after extensive discrimination training thus provides another instance of how selectivity in TE is shaped through experience.

4.3 LFP responses

4.3.1 Representation of occluded scenes by the LFP

LFPs are a measure of neuronal activity complementary to spike trains, as mentioned in Section 3.2.4. The latter reflect the activity of isolated pyramidal cells close to the electrode tip (Logothetis, 2002). In contrast, the LFP is a mass signal that is influenced by currents originating from axons, somata and dendrites around the electrode (Mitzdorf, 1987; Logothetis, 2002). While the behavior of single neurons in TE has been studied for several years, there has been far less effort devoted to the study of the LFP properties in area TE (Kreiman et al., 2004). Our study is the first in which stimulus effects on the LFP recorded in area TE are systematically studied.

The LFP reflects both ongoing cortical activity and stimulus evoked components (Mitzdorf, 1987). In a first analysis, the influence of a stimulus was separated from the ongoing activity by computing the VEPs. Because VEPs are derived by averaging repeated LFP measurements time-locked to the onset of the stimulus, any components without a fixed relationship to the stimulus are discarded. Presentation of natural scenes evoked three peaks in the VEPs of both monkeys, which were labeled N100, P130, and N200 to indicate their polarity and latency. Partial stimulus occlusion reduced the amplitudes of all three components, and increased their peak latencies. These effects systematically depended on how much of a stimulus was covered by an occluder. The more of an image was occluded, the stronger the amplitude reduction, and the longer the peak latency. Occlusion effects were similar across monkeys, although there was a consistent latency difference between the two animals.

In addition to general influences of occlusion, the VEP was also influenced by which image parts an occluder covered. Figures 3.40 and 3.41 show that in both monkeys there were sites at which the VEP amplitudes distinguished between diagnostic and non-diagnostic conditions. At these sites, the VEP amplitudes were only slightly influenced by the visible stimulus size of a condition. Similarly, there were sites in both monkeys where the VEP amplitudes were influenced mostly by the visible stimulus size, although the effects of the visible stimulus size were more pronounced for G00. These differences between occlusion conditions were observed for all three VEP components. In conclusion, the LFP recorded in area TE is sensitive to the placement of an occluder.

To analyze the effects more closely, individual recording sites were characterized by how strongly the two stimulus manipulations influenced the LFP amplitudes. As for the single units, the explained variance was computed for both factors. Since this measure characterized trial-to-trial variability of a signal, explained variances had to be computed from the raw LFP traces of each trial, and not from the VEP. Based on the explained variances, a clear influence of recording position on the behavior of LFP sites could be identified. The more anterior a recording site was located, the stronger the modulation of the LFP amplitude with stimulus diagnosticity. Also, the LFP sites with particularly strong influences of diagnosticity clustered anteriorly in both animals. This trend was present within each monkey. Furthermore, the data agreed well between monkeys. In contrast, the influence of visible stimulus size was homogeneous across the recording area.

The current data therefore shows a distinction between LFP and spikes. For the LFP, the influence of diagnosticity progressively increased from posterior to anterior locations. Yet, the single units in the whole area homogeneously encoded diagnosticity. To explain this difference, the sources generating the LFP need to be studied more closely. The LFP reflects the postsynaptic potentials in a certain region around the electrode tip. Postsynaptic potentials are generated because of the fibres providing input into an area from other brain regions. However, slightly more than half of the axonal tree of a pyramidal neuron remains within an area, while the other half leaves the area and forms long-range connections (A. Schüz, personal communication). Postsynaptic potentials from the axon parts that remain within the area will generate local contributions to the LFP. Thus, the LFP reflects both the local processing within a brain region, as well as the input into this region.

Since our recordings were carried out in TE only, the local processing component of the LFP can be assumed to be constant. This is also justified because the behavior of neurons was homogeneous throughout the region covered by the recordings. However, the input component differs between posterior and anterior TE. Considering only the major projections (see Section 1.2.1), both regions receive input from TEO, and project to each other. However, posterior TE exclusively receives projections from area V4. In conclusion, the input into posterior TE is dominated more by the influence of earlier visual areas than the input into anterior TE. Because the local part of the LFP was assumed to be constant, differences in the LFP between posterior and anterior TE point to a distinct behavior of the earlier visual areas. Our data suggest that the input into posterior TE from earlier visual areas carries only very limited information about stimulus diagnosticity. Yet, spiking activity of the single neurons in posterior TE, which reflects the processing within this brain region, discriminated between diagnostic and non-diagnostic conditions. Taken to-

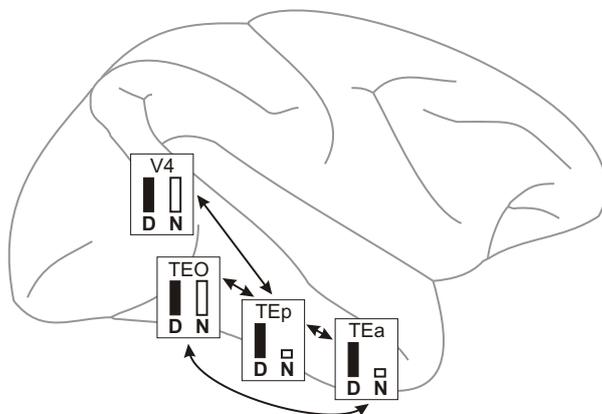


Figure 4.1: Model proposed based on the influences of diagnosticity on LFP and spike data. V_4 and TEO are included because they provide the main input to area TE, which is here subdivided into posterior TE (TEp) and anterior TE (TEa). The bar plots given for each area show hypothetical firing rates to diagnostic (black bar), and non-diagnostic (white bar) conditions. Arrows indicate connections between areas. As suggested in the text, neurons in V_4 and TEO do not distinguish between diagnostic and non-diagnostic conditions. These conditions are first discriminated in TEp.

gether, our results imply that the diagnosticity of an image feature is first encoded in posterior TE. Earlier visual areas like V_4 or TEO do not encode this stimulus dimension.

Neurons in V_4 respond to object features of intermediate complexity, resembling parts of an object boundary (Pasupathy & Connor, 1999, 2002). Similarly, TEO also responds to complex shapes (Kobatake & Tanaka, 1994). It has not been tested so far whether the encoding of these shape parts is influenced by their diagnosticity. Our data suggest that this is not the case. V_4 and TEO most likely provide TE with a description of shape parts; the processing in TE then assigns weights to these parts according to their diagnosticity.

The general model is outlined in Figure 4.1. Not only does it integrate our LFP and spike data, it also agrees with the reported effects of lesions in area TE. If, indeed, earlier visual areas do not extract diagnosticity, lesions of area TE have to result in the observed impairments in any task taxing the ability to identify diagnostic stimulus features.

The explained variances have been used to characterize spatial dependencies of the LFP and spike behavior. In both cases, data were not corrected for the ongoing or

spontaneous activity. This could not be avoided, because the random fluctuations present in the LFP can currently not be modeled. Consequently, their contributions could not be removed from the individual trial data. However, it is unlikely that differences in the spontaneous activity can account for the distinctly different spatial dependencies of LFP and spike data. Larger levels of noise lead to lower explained variance values. Different levels of spontaneous activity may thus explain why explained variances were generally lower for the LFP than for the spikes. However, to explain the spatial patterns, one has to assume that the spontaneous activity systematically decreases with more anterior recording locations only for the LFP. This already seems improbable at the outset. It furthermore predicts that for the LFP, the explained variances for the visible stimulus size should similarly increase from posterior to anterior locations. The reverse pattern was obtained experimentally, dismissing the possibility of artifacts because of the spontaneous activity.

Finally, the dependence of the LFP behavior on the recording location can explain the differences in the grand average VEPs from both monkeys. As mentioned above, there were sites in both monkeys at which the VEP amplitudes distinguished only between the diagnostic and non-diagnostic conditions, and the visible stimulus size had no effect. Nonetheless, the grand average VEP showed this effect only for monkey B98. For monkey G00, both diagnosticity and visible stimulus size influenced the VEPs equally strongly. The recording chamber was placed more posteriorly in monkey G00 than in monkey B98. Thus, the recordings from monkey G00 mostly sampled TE parts in which diagnosticity had only a small effect on the LFP amplitudes. In contrast, the more anteriorly located recording sites in monkey B98 showed stronger effects of diagnosticity on the LFP. This bias in sampling explains the differences in the grand average, which combined the data of all recording sites.

4.3.2 Comparison between the monkey LFP and the human EEG

Since no further LFP data is available for area TE, the observed effects can only be compared against EEG results from human observers. Similar to the LFP, the EEG captures the synchronized synaptic events of a neuronal population. However, the comparison between LFP and EEG is limited by a number of factors. The LFP reflects activity within a 0.5 - 3 mm radius around the electrode tip (Logothetis, 2002), with the electrode being placed in the brain area under study. In case of the EEG, electrodes are placed at the scalp, and therefore have a much larger distance from the cortical sources of the field potentials. Sources of EEG signals are thus most probably compact regions of cortex in which the local field activities of neurons

are synchronized on the cm^2 -scale. In addition, the generated local fields have to be similarly oriented by cortical geometry. Under these conditions, the interference of the fields of individual neurons generates a strong enough far-field EEG signal reaching the scalp by volume conduction (see e.g. Makeig et al., 2004). In addition to the EEG sources themselves being larger, volume conduction from other brain sources, as well as non-brain sources like muscles also influences the EEG signal recorded at an electrode. It is therefore rather difficult to attribute a recorded EEG component to a particular brain region (Schroeder et al., 1998). In any comparison between LFP and EEG, it has therefore to be kept in mind that the sources generating a signal change may be very different between the two signals. In addition, the polarity of a recorded signal depends on the positions of recording and reference electrode in the electromagnetic field. Because of the different placement of LFP and EEG electrodes, the same signal components may appear in LFP and EEG with different polarities. Finally, monkey and human brains have different sizes. It is therefore possible that the same components have longer latencies for humans than for monkeys.

The study by Doniger et al. (2000) described in Section 1.1.1 provides information on how occlusion affects the human VEP. Working with so called Snodgrass stimuli (Snodgrass & Feenan, 1990), Doniger et al. deleted varying amounts of complex objects. The study took advantage of the fact that objects can be recognized despite a certain level of deletion. The level of deletion at which objects can just be recognized is called the level of identification. If more of the object is deleted, subjects are no longer able to identify the object. VEPs were computed at the level of identification, and three larger deletion levels. Object presentation evoked a series of four VEP components: a positivity around 100 ms after stimulus onset (the P1), followed by a negativity around 170 ms (the N1), and a positivity at 220 ms (the P2). Finally, onsetting at the peak of the P2 and peaking at about 290 ms, there was a negative deflection in the VEPs (termed the N_{cl}). Both P1 and N1 were maximal at occipital sites, and were not influenced by stimulus deletion. The same was the case for the P2, which peaked at medial electrode sites (Doniger et al., 2001). In contrast, the amplitude, and judging from the figures in the paper, the latency of the N_{cl} systematically changed with occlusion. The less of a stimulus was deleted, the larger was the amplitude of the N_{cl} , and the shorter the latency. In addition to this gradual change with deletion level, there was a disproportionately large increase in the N_{cl} amplitude when the identification level was reached. Interestingly, the N_{cl} peaked at occipito-temporal scalp positions, suggesting that this component could be generated by activity in the LOC.

In the monkey grand average VEPs, the most consistent effects of occlusion were seen about 200 ms after stimulus onset. No influences of occlusion were present in the human VEPs at this time. Human VEPs began to differ between occlusion conditions about 30 ms later; the maximal occlusion effects were not reached until 90 ms after they reached their maximum in the monkey. As mentioned above, the latency of human EEG components could lag the latency of the corresponding monkey components. A factor of 5/3 between human and monkey latencies has been suggested to account for these differences (Doniger et al., 2000). The component corresponding to the monkey N200 therefore should have a latency of 330 ms in the human VEP, which is slightly longer than the actually observed latency of the N_{cl} . If the assumption of longer latencies in humans than in monkeys is correct, the N_{cl} in the human VEP may tentatively be identified with the N200 seen in the monkey VEP during recordings in area TE. A source for the N_{cl} in the LOC is consistent with this hypothesis.

With respect to earlier components, consistency between human and monkey VEP is less clear. For the N100 and P130, the monkey VEPs showed effects of occlusion. In the human VEP, no effects of occlusion were observed for the components preceding the N_{cl} . However, P1, N1 and P2 had different topographies than the N_{cl} , which suggests that these components were most likely generated from different sources. Indeed, both P1 and N1 are usually taken to reflect endogenous processing of visual stimuli (for summaries of the properties of P1 and N1 see Itier & Taylor, 2004a; Carmel & Bentin, 2002). The P1 is interpreted to reflect attention in spatial and non-spatial visual tasks. Its sources are estimated to lie in areas V3, V3a, and V4 for the early part of the component, and the fusiform gyrus for the late part (Di Russo et al., 2001). The N1 amplitude is larger if subjects perform a discrimination task than if they simply have to detect a stimulus. It is therefore taken to reflect a general discrimination process applied to a specific spatial location (Vogel & Luck, 2000). Source localizations show this component also to be generated from occipital brain regions (Itier & Taylor, 2004a,b). There are less studies on the P2 with respect to visual processes. However, it seems that the P2 recorded over the occipital cortex could reflect the processing of global figural geometry (Oka et al., 2001). As mentioned above, the P2 seems also to be generated by sources different than the sources for the N_{cl} . In conclusion, the VEP seems to be dominated by the influence of earlier visual areas before the onset of the N_{cl} . P1 and N1 thus most likely capture the behavior of brain regions whose activity does not affect the monkey VEPs measured directly in area TE. A comparison between the earlier components in the human VEP and the earlier components in the monkey data is thus not justified.

While the influence of categorization tasks on the human VEP has been studied, there is almost nothing known about which VEP responses are evoked by the diagnostic features themselves. The only available data come from an experiment by Smith et al. (2004), in which the Bubbles paradigm was combined with EEG recordings. Human observers were asked to classify faces based on their gender or expression. During the task, faces were only partially revealed by the usual random Bubbles masks. Behavioral responses were registered for every trial; at the same time, the EEG was recorded during the stimulus presentation. Sorting the Bubbles masks according to the behavioral responses revealed that – in agreement with the previous results by Gosselin & Schyns (2001) – different face regions were diagnostic for the two tasks. Whereas the eye region indicated the gender of a face, the mouth region was used to determine the expression. Bubbles masks could also be sorted based on the raw EEG amplitudes from each trial, to quantify the influence of the visibility of different face regions on the EEG. Bubbles masks from trials in which the EEG amplitudes at a particular time point were above the mean were contrasted with the masks from trials with amplitudes below the mean. The analysis was restricted to two components in the VEP, the P300 recorded over the parietal cortex, and the N170 recorded over occipito-temporal sites. The visibility of specific face regions correlated well with the amplitude of the EEG during these time periods. During the P300, the EEG amplitude was modulated by the presence of the diagnostic features, following the changes in diagnostic regions with task. In contrast, during the N170, EEG amplitudes were only influenced by the presence or absence of the eye region, irrespective of the task that the subjects performed. In contrast to the P300, the N170 seems therefore not to be influenced by stimulus diagnosticity.

The experiment of Smith et al. (2004) shows that in principle the contributions of diagnostic features to a certain VEP component can be quantified. Yet, it is difficult to relate the findings to our data. First of all, because of the recording location, only the N170 is relevant in this context. The N170 is taken to be a face-specific component, since its amplitude is larger for faces than for any other object category (Bentin et al., 1996; for a review, see e.g. Rousselet et al., 2004). The N170 sources are most likely found in a lateral temporal region near the superior temporal sulcus (Itier & Taylor, 2004b), a region also showing activation in fMRI studies on face perception and recognition (Haxby et al., 2000). Whether the N170 is specific to human faces or to faces in general, and how expertise with an object category contributes to this component is a matter of current debate (e.g. Rousselet et al., 2004; Rossion et al., 2004). Although animal faces were present in almost all images in our experiment, they covered only small portions of the scenes. In contrast, any study on the N170 has used close-up views of faces. Because of its face specificity, it is unclear whether the human N170 could even be elicited by the presentation of

our images. A comparison between the behavior of the human N170 and any of the monkey VEP components thus seems very far-fetched.

No further data are available on how the occlusion of specific image regions influences the human VEP. However, some conclusions about how the successful recovery of a partially occluded object affects the human VEP can be drawn from the two experiments by Doniger et al. (Doniger et al., 2000, 2001). Since the visibility of diagnostic features determines whether or not an image can be recognized, this comparison may give further insights into possible similarities between human and monkey VEPs. In both studies, the N_{cl} amplitudes followed the recognition performance of observers: At a fixed deletion level, the N_{cl} amplitudes were more negative when an object was identified than when identification failed. Furthermore, after repeated presentation of the same object, observers could identify the objects despite much higher amounts of deletion. The N_{cl} amplitude at a large level of deletion was more negative for the repeated object presentation than for the initial presentation, during which the object could not be identified. In conclusion, the N_{cl} amplitude indicates whether or not a partially visible object contour can successfully be completed into the full object. In parallel, the differences in the monkey VEP with diagnosticity implicate that the N200 amplitudes follow whether or not an image can be recognized despite the occlusion.

In conclusion, some similarities were observed between human and monkey VEP. These regard the N_{cl} in the human VEP, and the N200 in the monkey VEP. However, any conclusions are severely limited by the difficulty to precisely determine the sources of the scalp EEG. The topic thus requires further research. As a first step, the scalp EEG could be measured for the monkeys to allow more direct comparisons between LFP and EEG data.

4.4 Comparison of single unit and LFP responses

Next to the analysis of general differences, LFP and single unit behavior can be compared at a much finer scale. These comparisons take advantage of the fact that both signals were recorded with the same electrode. In agreement with the findings of Kreiman et al. (2004), little consistency was observed between corresponding LFP and spike data. LFP sites were in general more weakly influenced by changes in diagnosticity and visible stimulus size than the single units. The same was the case for stimulus selectivity, as long as it was quantified in terms of the explained variance. These effects can be explained by the fact that the LFP pools signals over a relatively large number of neurons. Diversity in this neuronal pool will reduce the net result of the different stimulus manipulations. Also, the LFP depends on the geometry of the sources generating the local fields, and therefore on the geometry of axons and dendrites. The LFP signal thus is not a simple average of the activity of all neurons in a region. Instead, local fields of individual neurons interfere depending on their orientation. Again, this can reduce effects seen in the LFP. A further explanation is also possible. The explained variances include random influences of the spontaneous activity. If these random fluctuations are higher in the LFP, the explained variances will also be lower for the LFP than for the single units.

While differences in the level of selectivity probably have to be expected when comparing LFP and spike data, it is more striking that LFP and neurons recorded at the same electrode respond maximally to completely different stimuli. Similarly, at a LFP recording site influenced strongly by diagnosticity, the chances of recording a neuron also strongly influenced by diagnosticity were not higher than at other recording sites. The same held true for the visible stimulus size. It therefore seems that the functional properties of the LFP and single units are quite distinct.

Since the LFP pools responses across a larger spatial extent, the fact that any selectivity could be observed at all implies a somewhat systematic topographical arrangement in the sources generating the LFP. Combining the LFP and MUA data, Kreiman et al. (2004) suggest that this conclusion concerns area TE. A clustering of TE neurons with similar selectivity would agree with the proposed columnar organization of area TE (Fujita et al., 1992; Wang et al., 1998). However, this hypothesis is inconsistent with our data. If the LFP behavior truly captures the properties of a group of neighboring TE neurons, then the behavior of individual neurons from this group should at least partially mimic the LFP behavior. In contrast, we consistently found no correlation between LFP and single cell behavior recorded at the same electrode. Since a large pool of neurons was tested, and because sometimes multiple neurons were recorded from the same electrode, it seems very

unlikely that these effects were simply generated due to chance sampling of only non-representative single units. Most likely, the TE spiking output therefore does not represent the sources of the LFP signal. Instead, it seems more plausible to identify the LFP with the input into TE, as has been assumed before. In this context, our data suggest that input into TE arrives in an orderly fashion, in the sense that the LFP arises from neuronal pools with relatively homogeneous properties. This seems plausible, since both V4 and TEO are still organized in a retinotopic manner (Gattass et al., 1988; Boussaoud et al., 1991). Furthermore, the foveal representation of V4 contains group of cells with similar orientation preferences, suggesting that some functional organization may exist in V4 (Ghose & Ts'o, 1997). At the same time, our results highlight that the input that a neuron TE receives does not necessarily predict the functional behavior of the neuron in a straightforward manner.

4.5 Further experiments

A number of interesting questions arise from the findings of this projects, which could be pursued in further experiments. Behaviorally, monkeys found other image regions diagnostic than human observers. Especially puzzling was the finding that monkey B98 relied more on background regions than on object parts. In our experiments the monkeys only had to discriminate among three images, which were always presented at the center of the screen. Because of the positional invariance and the limited stimulus set, in addition to the complex backgrounds of natural scenes, almost any image region contained sufficient differences between the three images to be used to discriminate between them. In human observers, the bias to use object features seems to have been strong enough to still rely on object regions. However, this bias seems not to be as pronounced in monkeys. It has been demonstrated that monkeys need additional processing time to perform a categorization task in comparison to a simple stimulus detection task (Macé et al., 2005). This could be an indication that they have to extract additional, possibly more general features in the categorization task. However, the results by D'Amato & van Sant (1988) show that this does not necessarily have to be the case. As a first additional experiment, it would be interesting to test which object features become diagnostic when monkeys have to perform a more complex task than simple image identification. One possibility is to test them on a categorization task. Another possibility is to use the same identification task as before, but to manipulate the members of the stimulus set. Images for example could be shown at different screen positions, or they could appear in several rotations. Based on the outcomes of these experiments, further conclusions about Rhesus monkeys' ability to generalize can be drawn.

A second experiment, which would be interesting both behaviorally and neurophysiologically, follows along the same lines. It has been argued several times that diagnostic features are determined by the task and the stimulus set. The described experiments by Gosselin and Schyns indeed demonstrate that one stimulus set can have different diagnostic regions, depending on the task. Behaviorally, determining diagnostic regions for the same stimuli under different task constraints provides another instance of testing the monkeys' ability to generalize. On the neurophysiological level, the fact that different features of a stimulus are diagnostic for different tasks makes it possible to disambiguate the encoding of a certain feature from the encoding of its diagnosticity. It can be tested whether TE neurons always respond to the presence of a certain feature, or whether they follow the diagnosticity of the features, and thus truly indicate diagnosticity.

The following experiment is therefore proposed: Monkeys will be trained to perform two discrimination tasks on one fixed image set. Because of the success with human and pigeons, a good option would be to train them to discriminate faces either based on the gender or the expression. Using Bubbles, the diagnostic regions of the faces will then be determined for each of the tasks. If different regions are used to determine the gender or expression, these results can be used to construct stimuli for a subsequent neurophysiological study. As in our experiment, stimuli will be constructed by masking the faces, so that either the diagnostic regions of the gender task, or the diagnostic regions of the face task remain visible. Neurons will be recorded in area TE, while the monkey actively performs the two tasks. Recording sessions will start with one of the tasks; after sufficient data collection, the monkeys will be switched to the other. To test how TE neurons follow the changes in diagnosticity across tasks, it is necessary to always record neural responses to the diagnostic regions of both tasks. However, it is not straightforward to collect this data because of the following reason: Assume that the monkey performs the gender task. Presentation of full faces will elicit correct responses, as will the presentation of the diagnostic regions for the gender task. Yet, ideally the diagnostic regions for the expression task are not diagnostic for the gender task, and the monkey cannot use these regions to correctly perform the task. This introduces differences in the recognition rates that can be expected for the different conditions, and poses problems with respect to rewarding the monkey. The simplest solution to the problem seems to be as follows. While the monkey performs a task, trials will be randomly interspersed in which no response is required, and simple fixation on the stimulus is sufficient for reward delivery. These trials could be indicated to the monkey e.g. by a change in the color of the fixation spot. In our present experiment, the selectivity shaped by training proved sufficient to detect differences between diagnostic and non-diagnostic conditions even without the context of an identification task. For the suggested experiment, it is assumed that the performance of a task provides the necessary framework and adjusts the neural selectivity accordingly, which should also be seen in the fixation trials. The monkey will be asked to perform the task only on the full faces and the diagnostic regions for the current task. Fixation trials will contain the full faces and the task-matched diagnostic regions for comparison, in addition to the diagnostic regions of the other task. This however will result in the task-matched diagnostic regions to be shown more often. To mask these differences, a certain proportion of the task trials could contain faces masked with the standard Bubbles masks. Comparing the responses to full faces and the task-matched diagnostic regions between task and fixation trials can later serve to detect any possible differences due to the different trial types. If these are not present, responses to the diagnostic regions of the two tasks can safely be compared. In conclusion, it can

be tested whether TE neurons respond to a particular face region irrespective of its diagnosticity, or, alternatively, whether the diagnosticity determines the responses to certain features.

Finally, it was suggested that the encoding of diagnosticity is a property which is first achieved in area TE. The hypothesis is consistent with lesion data. In addition, a modulation of neural firing rate with diagnosticity is consistent with previous studies in area TE. However, it has not been tested so far whether neurons in areas V4 and TEO can distinguish between diagnostic and non-diagnostic conditions. In our project, the conclusions about the properties of these areas were drawn from the LFP data. To confirm the model, single cell recordings should be carried out directly in area V4 and TEO. If the assumptions prove to be correct, neurons in these areas will not respond differently to diagnostic and non-diagnostic conditions. It can then safely be concluded that area TE does not only encode diagnosticity, but that one function of this area is to extract the diagnosticity of features. Simultaneous recordings in areas V4, TEO, and TE could also be used to address another issue. Our data showed that the LFP behavior recorded at an electrode did not correlate well with the behavior of the single neurons recorded at the same electrode. It was suggested that this effect could be explained by the assumption that the LFP behavior reflects the selectivity of the areas projecting to TE, but not of TE itself. The simultaneous recordings could address this question. If the hypothesis is correct, than the spiking activity in V4 or TEO should be a better predictor of the LFP recorded in area TE than the spiking activity of TE neurons.

Chapter 5

Conclusions

Monkeys were trained to discriminate between natural scenes. Covering images with randomly constructed occluders during the discrimination task revealed that the occlusion of certain image regions interacted with the monkeys' recognition performance. Each natural scene could thus be subdivided into a diagnostic and a non-diagnostic region. The diagnostic regions were characteristic for each monkey. They could not be predicted from the behavior of human observers, nor from the physical characteristics of different image regions. Instead, they reflected the observers' particular strategy and biases. The data thus show that it can be misleading to infer monkey behavior from human behavior, as is commonly done. This is the case even if according to standard measures like the percentage of correct trials, the monkeys seem to perform similarly to human observers.

The behavioral data was subsequently used to test how the placement of an occluder influences the responses of TE neurons. Two parameters were used to construct different occluders: Occluders could be placed to cover diagnostic or non-diagnostic image regions (factor: diagnosticity), and they could have one of three sizes (factor: visible stimulus size). The two factors had different behavioral consequences for the monkeys: Recognition performance was unimpaired despite occlusion as long as the diagnostic image parts were visible. In contrast to this pronounced effect of diagnosticity, the visible stimulus size had little effect on recognition rates. The responses of TE neurons systematically varied with occluder placement. Large differences were observed between diagnostic and non-diagnostic conditions. In general, the neurons were more responsive and more selective in the diagnostic conditions. In contrast, the visible stimulus size had only little impact on the neural behavior. The effects of occlusion on TE responses therefore followed the behavioral consequences of the occluder. As long as an image could be identified despite the occlusion, the

neurons remained responsive. With respect to the responses evoked by unoccluded images, occlusion led to a general response reduction. This reduction was however significantly larger for the non-diagnostic than for the diagnostic conditions. In conclusion, these experiments reveal how different parts of natural scenes contribute to the neural responses to the full scene. The results suggest that the impact of an image region on the firing rate in TE is determined by the diagnosticity of the region.

Together with the spiking activity of individual neurons, the LFP was recorded at each electrode. Stimulus-locked components in the LFP were characterized by computing the VEPs. Analysis was performed on three major VEP components, the N100, P130, and N200. Occlusion had effects on all three components in both monkeys. LFP sites could be identified where VEP amplitudes distinguished among diagnostic and non-diagnostic conditions (diagnosticity sites). The visible stimulus size had little effect on the VEP amplitudes at these sites. Similarly, there were sites where the VEP amplitude was mostly influenced by the visible stimulus size (size sites). Both types of LFP sites were present in both monkeys. The size sites were distributed homogeneously throughout the TE region covered during the recording. In contrast, the diagnosticity sites were localized mainly in anterior TE. This trend could also be seen when analyzing the modulations in raw LFP amplitudes introduced by changes in diagnosticity, or changes in the visible stimulus size. Again, the influence of diagnosticity on the LFP was stronger in the anterior part of TE than in the posterior part. This was the case for each monkey individually, as well as for the combined data. For the spike data, there was no such dependency on the recording locations. The LFP reflects local processing, as well as the input of other visual areas into the studied brain region. The behavior of the single units indicates that the local processing is homogeneous throughout TE. Because of the constant local processing, differences in the LFP across TE thus have to be due to differences in the input into the anterior and posterior part. Indeed, anterior and posterior TE receive inputs from different brain regions. Most notably, the projections from area V4 terminate in posterior TE exclusively. Since diagnosticity did not exert strong effects on the LFP amplitudes recorded in posterior TE, our results imply that the earlier visual areas exclusively projecting to posterior TE are not sensitive to the diagnosticity of a stimulus feature. Based on the observed behavior of the single neurons, diagnosticity is clearly encoded in the output of posterior TE. Our data therefore suggest that stimulus diagnosticity is first encoded in posterior TE, but not in the earlier visual areas. The model not only explains the dependency of the LFP behavior on recording location. It is also consistent with the effects of lesions in area TE, which impair tasks taxing the ability to extract diagnostic information from visual input. The LFP results therefore allow to extend

the conclusions drawn from the single unit behavior: It seems that neurons in area TE do not just encode diagnosticity, but that it is one function of area TE to extract this information.

Bibliography

- Albright, T. D. & Gross, C. G. (1990). Do inferior temporal cortex neurons encode shape by acting as Fourier Descriptor filters? In *Proceedings of the International Conference on Fuzzy Logic & Neural Networks*. Izuka, Japan, pp. 375–378.
- Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, *103* (1), 62–70.
- Baker, C. I., Behrmann, M. & Olson, C. R. (2002). Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nature Neuroscience*, *5* (11), 1210–1216.
- Bakin, J. S., Nakayama, K. & Gilbert, C. D. (2000). Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *Journal of Neuroscience*, *20* (21), 8188–8198.
- Baylis, G. C. & Driver, J. (2001). Shape-coding in IT cells generalizes over contrast and mirror-reversal, but not figure-ground reversal. *Nature Neuroscience*, *4* (9), 937–942.
- Baylis, G. C., Rolls, E. T. & Leonard, C. M. (1987). Functional Subdivisions of the Temporal Lobe Neocortex. *Journal of Neuroscience*, *7* (2), 330–342.
- Bentin, S., Allison, T., Puce, A., Perez, E. & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8* (6), 551–565.
- Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, *94* (2), 115–147.
- Blough, D. S. (1984). Form recognition in pigeons. In H. L. Roitblat, T. G. Bever & H. S. Terrace (eds.), *Animal cognition*. Hillsdale, NJ: Erlbaum, pp. 277–289.
- Bortz, J. (1993). *Statistik für Sozialwissenschaftler*. Berlin: Springer Verlag.

- Boussaoud, D., Desimone, R. & Ungerleider, L. G. (1991). Visual topography of area TEO in the macaque. *Journal of Comparative Neurology*, *306*, 554–575.
- Bruno, N., Bertamini, M. & Domini, F. (1997). Amodal completion of partly occluded surfaces: Is there a mosaic stage? *Journal of Experimental Psychology: Human Perception & Performance*, *23* (5), 1412–1426.
- Buffalo, E. A., Bertini, G., Ungerleider, L. G. & Desimone, R. (2005). Impaired filtering of distracter stimuli by TE neurons following V4 and TEO lesions in macaques. *Cerebral Cortex*, *15*, 141–151.
- Carmel, D. & Bentin, S. (2002). Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition*, *83*, 1–29.
- Cerella, J. (1980). The pigeon's analysis of pictures. *Pattern Recognition*, *12*, 1–6.
- Chauvin, A., Worsley, K. J., Schyns, P. G., Arguin, M. & Gosselin, F. (submitted). A sensitive statistical test for smooth classification images.
- Cheng, K., Saleem, K. S. & Tanaka, K. (1997). Organization of corticostriatal and corticoamygdalar projections arising from the anterior inferotemporal area TE of the macaque monkey: A Phaseolus vulgaris leucoagglutinin study. *Journal of Neuroscience*, *17* (20), 7902–7925.
- Craton, L. G. (1996). The development of perceptual completion abilities: Infants' perception of stationary, partially occluded objects. *Child Development*, *67*, 890–904.
- D'Amato, M. R. & van Sant, P. (1988). The person concept in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes*, *14* (1), 43–55.
- De Weerd, P., Desimone, R. & Ungerleider, L. G. (2003). Generalized deficits in visual selective attention after V4 and TEO lesions in macaques. *European Journal of Neuroscience*, *18*, 1671–1691.
- Dean, P. (1976). Effects of inferotemporal lesions on the behavior of monkeys. *Psychological Bulletin*, *83* (1), 41–71.
- Dean, P. (1982). Visual behavior in monkeys with inferotemporal lesions. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (eds.), *Analysis of visual behavior*. Cambridge, MA: MIT Press, pp. 587–628.

- Deruelle, C., Barbet, I., Dépy, D. & Fagot, J. (2000). Perception of partly occluded figures by baboons (*Papio papio*). *Perception*, *29*, 1483–1497.
- Desimone, R., Albright, T. D., Gross, C. G. & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4* (8), 2051–2062.
- Desimone, R., Fleming, J. F. R. & Gross, C. G. (1980). Prestriate afferents to inferior temporal cortex: An HRP study. *Brain Research*, *184*, 41–55.
- Desimone, R. & Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Research*, *178* (2-3), 363–380.
- Di Russo, F., Martínez, A., Sereno, M. I., Pitzalis, S. & Hillyard, S. A. (2001). Cortical sources of early components of the visual evoked potential. *Human Brain Mapping*, *15*, 95–111.
- DiCarlo, J. J. & Maunsell, J. H. R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology*, *89*, 3264–3278.
- DiPietro, N. T., Wasserman, E. A. & Young, M. E. (2002). Effects of occlusion on pigeons' visual object recognition. *Perception*, *31*, 1299–1312.
- Distler, C., Boussaoud, D., Desimone, R. & Ungerleider, L. G. (1993). Cortical connections of inferior temporal area TEO in the macaque monkey. *Journal of Comparative Neurology*, *334*, 125–150.
- Doniger, G. M., Foxe, J. J., Murray, M. M., Higgins, B. A., Snodgrass, J. G., Schroeder, C. E. & Javitt, D. C. (2000). Activation timecourse of ventral visual stream object-recognition areas: High density electrical mapping of perceptual closure processes. *Journal of Cognitive Neuroscience*, *12* (4), 615–621.
- Doniger, G. M., Foxe, J. J., Schroeder, C. E., Murray, M. M., Higgins, B. A. & Javitt, D. C. (2001). Visual perceptual learning in human object recognition areas: A repetition priming study using high-density electrical mapping. *Neuroimage*, *13* (2), 305–313.
- Elston, G. N. & Rosa, M. G. (1998). Morphological variation of layer III pyramidal neurones in the occipitotemporal pathway of the macaque monkey visual cortex. *Cerebral Cortex*, *8* (3), 278–294.

- Elston, G. N., Tweedale, R. & Rosa, M. G. (1999). Cortical integration in the visual system of the macaque monkey: Large-scale morphological differences in the pyramidal neurons in the occipital, parietal and temporal lobes. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 266 (1426), 1367–1374.
- Felleman, D. J. & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1 (1), 1–47.
- Foley, M. A., Foley, H. J., Durso, F. T. & Smith, N. K. (1997). Investigations of closure processes: What source-monitoring judgments suggest about what is "closing". *Memory & Cognition*, 25 (2), 140–155.
- Forkman, B. (1998). Hens use occlusion to judge depth in a two-dimensional picture. *Perception*, 27, 861–867.
- Fujita, I., Tanaka, K., Ito, M. & Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex. *Nature*, 360, 343–346.
- Fujita, K. (2001a). Perceptual completion in rhesus monkeys (*Macaca mulatta*) and pigeons (*Columbia livia*). *Perception & Psychophysics*, 63 (1), 115–125.
- Fujita, K. (2001b). What you see is different from what I see: Species differences in visual perception. In T. Matsuzawa (ed.), *Primate origins of human cognition and behavior*. Tokyo: Springer, pp. 29–54.
- Gaffan, E. A., Harrison, S. & Gaffan, D. (1986). Single and concurrent discrimination learning by monkeys after lesions of inferotemporal cortex. *The Quarterly Journal of Experimental Psychology*, 38B, 31–51.
- Gattass, R., Sousa, A. P. B. & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *Journal of Neuroscience*, 8 (6), 1831–1845.
- Ghose, G. M. & Ts'o, D. (1997). Form processing modules in primate area V4. *Journal of Neurophysiology*, 77, 2191–2196.
- Gibson, B. M., Wasserman, E. A., Gosselin, F. & Schyns, P. G. (in press). Pigeons and people categorize human faces using similar features. *Journal of Experimental Psychology: Animal Behavior Processes*.
- Gochin, P. M. (1996). The representation of shape in the temporal lobe. *Behavioural Brain Research*, 76, 99–116.

- Gold, J. M., Murray, R. F., Bennett, P. J. & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology*, *10* (11), 663–666.
- Gosselin, F. & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision Research*, *41*, 2261–2271.
- Gosselin, F. & Schyns, P. G. (2005). Bubbles: A user's guide. In L. Gershkoff-Stowe & D. H. Rakison (eds.), *Building Object Categories in Developmental Time*. Hillsdale, NJ: Erlbaum, pp. 91–106.
- Greene, S. L. (1983). Feature memorization in pigeon concept formation. In M. L. Commons, R. J. Herrnstein & A. R. Wagner (eds.), *Quantitative analyses of behavior: Discrimination processes*. Cambridge, MA: Ballinger, pp. 209–229.
- Gross, C. G., Bender, D. B. & Gerstein, G. L. (1979). Activity of inferior temporal neurons in behaving monkeys. *Neuropsychologia*, *17*, 215–229.
- Gross, C. G., Bender, D. B. & Mishkin, M. (1977). Contributions of the corpus callosum and the anterior commissure to visual activation of inferior temporal neurons. *Brain Research*, *131*, 227–239.
- Gross, C. G., Bender, D. B. & Rocha-Miranda, C. E. (1969). Visual receptive fields of neurons in inferotemporal cortex of the monkey. *Science*, *166*, 1303–1306.
- Gross, C. G., Rocha-Miranda, C. E. & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the macaque. *Journal of Neurophysiology*, *35*, 96–111.
- Guo, K., Robertson, R. G., Mahmoodi, S., Tadmor, Y. & Young, M. P. (2003). How do monkeys view faces? - A study of eye movements. *Experimental Brain Research*, *150*, 363–374.
- Hasselmo, M. E., Rolls, E. T. & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*, *32*, 203–218.
- Haxby, J. V., Hoffman, E. A. & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4* (6), 223–233.
- He, Z. J. & Nakayama, K. (1992). Surface features in visual search. *Nature*, *359*, 231–233.
- Helmholtz, H. v. (1910). *Treatise on physiological optics*. New York: Dover.

- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in Cognitive Sciences*, 7 (11), 498–504.
- Henderson, J. M. & Hollingworth, A. (1999). High-level scene perception. *Annual Reviews of Psychology*, 50, 243–271.
- Henderson, J. M., Weeks, P. A. J. & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception & Performance*, 25 (1), 210–228.
- Hikosaka, K. (1999). Tolerances of responses to visual patterns in neurons of the posterior inferotemporal cortex in the macaque against changing stimulus size and orientation, and deleting patterns. *Behavioural Brain Research*, 100, 67–76.
- Hubel, D. H. & Wiesel, T. N. (1977). Ferrier Lecture: Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, 98 (1130), 1–59.
- Huxlin, K. R., Saunders, R. C., Marchionini, D., Pham, H. A. & Merigan, W. H. (2000). Perceptual deficits after lesions of inferotemporal cortex in macaques. *Cerebral Cortex*, 10 (7), 671–683.
- Itier, R. J. & Taylor, M. J. (2004a). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex*, 14, 132–142.
- Itier, R. J. & Taylor, M. J. (2004b). Source analysis of the N170 to faces and objects. *NeuroReport*, 15 (8), 1261–1265.
- Itti, L., Koch, C. & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (11), 1254–1259.
- Iwai, E. (1985). Neuropsychological basis of pattern vision in macaque monkeys. *Vision Research*, 25, 425–439.
- Iwai, E. & Mishkin, M. (1969). Further evidence of the locus of the visual area in the temporal lobe of the monkey. *Experimental Neurology*, 25, 585–594.
- Iwai, E. & Yukie, M. (1987). Amygdalofugal and amygdalopetal connections with modality-specific visual cortical areas in macaques (*Macaca fuscata*, *M. mulatta*, and *M. fascicularis*). *Journal of Comparative Neurology*, 261 (3), 362–387.

- Iwai, E., Yukie, M., Watanabe, J., Yaginuma, S., Osawa, Y. & Hikosaka, K. (1990). Two-stage model of visual pattern discrimination learning in macaque monkeys (*Macaca mulatta* and *Macaca fuscata*). *Tohoku Journal of Experimental Medicine*, *161 Suppl.*, 79–93.
- Janssen, P., Vogels, R. & Orban, G. A. (2000). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science*, *288*, 2054–2056.
- Joseph, J. S. & Nakayama, K. (1999). Amodal representation depends on the object seen before partial occlusion. *Vision Research*, *39* (2), 283–292.
- Judge, S. J., Richmond, B. J. & Chu, F. C. (1980). Implantation of magnetic search coils for measurement of eye position: An improved method. *Vision Research*, *20*, 535–538.
- Kanizsa, G. & Gerbino, W. (1982). Amodal completion: Seeing or Thinking? In J. Beck (ed.), *Organization and representation in perception*. Hillsdale, NJ: Erlbaum, pp. 167–190.
- Kanizsa, G., Renzi, P., Conte, S., Compostela, C. & Guerani, L. (1993). Amodal completion in mouse vision. *Perception*, *22*, 713–721.
- Kanwisher, N., Chun, M. M., McDermott, J. & Ledden, P. J. (1996). Functional imaging of human visual recognition. *Cognitive Brain Research*, *5*, 55–67.
- Kellman, P. J. & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*, 483–524.
- Kikuchi, R. & Iwai, E. (1980). The locus of the posterior subdivision of the inferotemporal visual learning in the monkey. *Brain Research*, *198*, 347–360.
- Kobatake, E. & Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiology*, *71* (3), 856–867.
- Kobatake, E., Wang, G. & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology*, *80*, 324–330.
- Kourtzi, Z. & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science*, *293*, 1506–1509.
- Kovács, G., Vogels, R. & Orban, G. A. (1995). Selectivity of macaque inferior temporal neurons for partially occluded shapes. *Journal of Neuroscience*, *15* (3), 1984–1997.

- Kreiman, G., Hung, C., Shkolnik, A., Poggio, T. & DiCarlo, J. J. (2004). Object recognition by selective spike and LFP data in inferior temporal cortex. *Society for Neuroscience Abstracts*, 751.15.
- Lea, S. E. G., Slater, A. M. & Ryan, C. M. E. (1996). Perception of object unity in chicks: A comparison with the human infant. *Infant Behavior and Development*, 19, 501–504.
- Lee, A. (2001). Membrane structure. *Current Biology*, 11 (20), 811–814.
- Lerner, Y., Hendler, T. & Malach, R. (2002). Object-completion effects in the human lateral occipital complex. *Cerebral Cortex*, 12, 163–177.
- Loftus, G. R. & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception & Performance*, 4 (4), 565–572.
- Logothetis, N. K. (2002). The neural basis of the blood-oxygen-level-dependent functional magnetic resonance imaging signal. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 357, 1003–1037.
- Logothetis, N. K., Pauls, J. & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology*, 5 (5), 552–563.
- Logothetis, N. K. & Sheinberg, D. L. (1996). Visual object recognition. *Annual Review of Neuroscience*, 19, 577–621.
- Macé, M. J. M., Richard, G., Delorme, A. & Fabre-Thorpe, M. (2005). Rapid categorization of natural scenes in monkeys: target predictability and processing speed. *NeuroReport*, 16 (4), 349–354.
- Mackworth, N. H. & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics*, 2 (11), 547–552.
- Makeig, S., Debener, S., Onton, J. & Delorme, A. (2004). Mining event-related brain dynamics. *Trends in Cognitive Sciences*, 8 (5), 204–210.
- Mannan, S. K., Ruddock, K. H. & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision*, 9 (3), 363–386.
- Mannan, S. K., Ruddock, K. H. & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, 10 (3), 165–188.

- Missal, M., Vogels, R. & Orban, G. A. (1997). Responses of macaque inferior temporal neurons to overlapping shapes. *Cerebral Cortex*, *7*, 758–767.
- Mitzdorf, U. (1987). Properties of the evoked potential generators: Current source-density analysis of visually evoked potentials in the cat cortex. *International Journal of Neuroscience*, *33*, 33–59.
- Miyashita, Y., Okuno, H., Tokuyama, W., Ihara, T. & Nakajima, K. (1996). Feedback signal from medial temporal lobe mediates visual associative mnemonic codes of inferotemporal neurons. *Cognitive Brain Research*, *5* (1-2), 81–86.
- Morel, A. & Bullier, J. (1990). Anatomical segregation of two cortical visual pathways in the macaque monkey. *Visual Neuroscience*, *4*, 555–578.
- Murray, M. M., Foxe, D. M., Javitt, D. C. & Foxe, J. J. (2004). Setting boundaries: Brain dynamics of modal and amodal illusory shape completion in humans. *Journal of Neuroscience*, *24* (31), 6898–6903.
- Murray, R. F. & Gold, J. M. (2004). Troubles with bubbles. *Vision Research*, *44*, 461–470.
- Murray, R. F., Sekuler, A. B. & Bennett, P. J. (2001). Time course of amodal completion revealed by a shape discrimination task. *Psychonomic Bulletin & Review*, *8* (4), 713–720.
- Nahm, F. K. D., Perret, A., Amaral, D. G. & Albright, T. D. (1997). How do monkeys look at faces? *Journal of Cognitive Neuroscience*, *9* (5), 611–623.
- Oka, S., van Tonder, G. & Ejima, Y. (2001). A VEP study on visual processing of figural geometry. *Vision Research*, *41*, 3791–3803.
- Op de Beeck, H. & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology*, *426*, 505–518.
- Osada, Y. & Schiller, P. H. (1994). Can monkeys see objects under conditions of transparency and occlusion. *Investigative Ophthalmology & Visual Science*, *35* (4), 1664.
- Panzeri, S. & Treves, A. (1996). Analytical estimates of limited sampling biases in different information measures. *Network: Computation in Neural Systems*, *7*, 87–107.
- Pasupathy, A. & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, *82* (5), 2490–2502.

- Pasupathy, A. & Connor, C. E. (2002). Population coding of shape in area V4. *Nature Neuroscience*, 5 (12), 1332–1338.
- Perrett, D. I., Hietanen, J. K., Oram, M. W. & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, 335, 23–30.
- Perrett, D. I., Oram, M. W., Harries, M. H., Bevan, R., Hietanen, J. K., Benson, P. J. & Thomas, S. (1991). Viewer-centered and object-centered coding of heads in the macaque temporal cortex. *Experimental Brain Research*, 86, 159–173.
- Perrett, D. I., Rolls, E. T. & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, 47, 329–342.
- Rauschenberger, R. & Yantis, S. (2001). Masking unveils pre-amodal completion representation in visual search. *Nature*, 410, 369–372.
- Regolin, L. & Vallortigara, G. (1995). Perception of partly occluded objects by young chicks. *Perception & Psychophysics*, 57 (7), 971–976.
- Reinagel, P. & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems*, 10, 341–350.
- Richmond, B. J. & Optican, L. M. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. Quantification of response waveform. *Journal of Neurophysiology*, 57 (1), 147–161.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R. & Bialek, W. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: The MIT Press.
- Robinson, D. A. (1963). A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Transactions in Biomedical Engineering*, BME-10, 137–145.
- Ross, J., Morrone, M. C., Goldberg, M. E. & Burr, D. C. (2001). Changes of visual perception at the time of saccades. *Trends in Neurosciences*, 24 (2), 113–121.
- Rossion, B., Kung, C.-C. & Tarr, M. J. (2004). Visual expertise with nonface objects leads to competition with the early perceptual processing of faces in the human occipitotemporal cortex. *Proceedings of the National Academy of Sciences (USA)*, 101 (40), 14521–14526.

- Rousselet, G. A., Macé, M. J. M. & Fabre-Thorpe, M. (2004). Animal and human faces in natural scenes: How specific to human faces is the N170 ERP component? *Journal of Vision*, *4*, 13–21.
- Sakai, K. & Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. *Nature*, *354*, 152–155.
- Sakai, K. & Miyashita, Y. (1994). Neuronal tuning to learned complex forms in vision. *NeuroReport*, *5*, 829–832.
- Saleem, K. S., Suzuki, W., Tanaka, K. & Hashikawa, T. (2000). Connections between anterior inferotemporal cortex and superior temporal sulcus regions in the macaque monkey. *Journal of Neuroscience*, *20* (13), 5083–5101.
- Saleem, K. S. & Tanaka, K. (1996). Divergent projections from the anterior inferotemporal area TE to the perirhinal and entorhinal cortices in the macaque monkey. *Journal of Neuroscience*, *16* (15), 4757–4775.
- Sands, S. F., Lincoln, C. E. & Wright, A. A. (1982). Pictorial similarity judgments and the organization of visual memory in the Rhesus monkey. *Journal of Experimental Psychology: General*, *111* (4), 369–389.
- Schiller, P. H. & Koerner, F. (1971). Discharge characteristics of single units in superior colliculus of the alert rhesus monkey. *Journal of Neurophysiology*, *34*, 920–936.
- Schroeder, C. E., Mehta, A. D. & Givre, S. J. (1998). A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. *Cerebral Cortex*, *8*, 575–592.
- Schwartz, E. L., Desimone, R., Albright, T. D. & Gross, C. G. (1983). Shape recognition and inferior temporal neurons. *Proceedings of the National Academy of Sciences (USA)*, *80*, 5776–5778.
- Sekuler, A. B., Lee, J. A. J. & Shettleworth, S. J. (1996). Pigeons do not complete partly occluded figures. *Perception*, *25*, 1109–1120.
- Sekuler, A. B. & Palmer, S. E. (1992). Perception of partly occluded objects: a microgenetic analysis. *Journal of Experimental Psychology: General*, *121* (1), 95–111.
- Sekuler, A. B., Palmer, S. E. & Flynn, C. (1994). Local and global processes in visual completion. *Psychological Science*, *5* (5), 260–267.

- Seltzer, B. & Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Research*, *149*, 1–24.
- Shannon, C. E. & Weaver, W. (1949). *The mathematical theory of communication*. Urbana: University of Illinois Press.
- Sheinberg, D. L. & Logothetis, N. K. (2001). Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. *Journal of Neuroscience*, *21*, 1340–1350.
- Shimojo, S. & Nakayama, K. (1990). Amodal representation of occluded surfaces: Role of invisible stimuli in apparent motion correspondence. *Perception*, *19* (3), 285–299.
- Shiwa, T. (1987). Corticocortical projections to the monkey temporal lobe with particular reference to the visual processing pathways. *Archives Italiennes de Biologie*, *125* (2), 139–154.
- Shore, D. I. & Enns, J. T. (1997). Shape completion time depends on the size of the occluded region. *Journal of Experimental Psychology: Human Perception & Performance*, *23* (4), 980–998.
- Siegel, S. (2001). *Nichtparametrische statistische Methoden*. Eschborn: Verlag Dietmar Klotz.
- Sigala, N., Gabbiani, F. & Logothetis, N. K. (2002). Visual categorization and object representation in monkeys and humans. *Journal of Cognitive Neuroscience*, *14* (2), 187–198.
- Sigala, N. & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*, 318–320.
- Smith, M. L., Gosselin, F. & Schyns, P. G. (2004). Receptive fields for flexible face categorizations. *Psychological Science*, *15* (11), 753–761.
- Snodgrass, J. G. & Feenan, K. (1990). Priming effects in picture fragment completion: Support for the perceptual closure hypothesis. *Journal of Experimental Psychology: General*, *119* (3), 276–296.
- Spitzer, H. & Richmond, B. J. (1991). Task difficulty: ignoring, attending to, and discriminating a visual stimulus yield progressively more activity in inferior temporal neurons. *Experimental Brain Research*, *83*, 340–348.

- Squire, L. R., Stark, C. E. L. & Clark, R. E. (2004). The medial temporal lobe. *Annual Review of Neuroscience*, *27*, 279–306.
- Stevens, K. A. (1983). Evidence relating subjective contours and interpretations involving interposition. *Perception*, *12*, 491–500.
- Sugita, Y. (1999). Grouping of image fragments in primary visual cortex. *Nature*, *401*, 269–272.
- Suzuki, W., Saleem, K. S. & Tanaka, K. (2000). Divergent backward projections from the anterior part of the inferotemporal cortex (area TE) in the macaque. *Journal of Comparative Neurology*, *422* (2), 206–228.
- Tamura, H. & Tanaka, K. (2001). Visual response properties of cells in the ventral and dorsal parts of the macaque inferotemporal cortex. *Cerebral Cortex*, *11* (5), 384–399.
- Tanaka, K. (1993). Neuronal mechanisms of object recognition. *Science*, *262*, 685–688.
- Tanaka, K. (1996). Inferotemporal cortex and object vision. *Annual Review of Neuroscience*, *19*, 109–139.
- Tanaka, K. (2000). Mechanisms of visual object recognition studied in monkeys. *Spatial Vision*, *13* (2,3), 147–163.
- Tanaka, K., Saito, H. A., Fukada, Y. & Moriya, M. (1991). Coding visual images of objects in the inferotemporal cortex of the macaque monkey. *Journal of Neurophysiology*, *66*, 170–189.
- Tsunoda, K., Yamane, Y., Nishizaki, M. & Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, *4* (8), 832–838.
- Ungerleider, L. G., Gaffan, D. & Pelak, V. S. (1989). Projections from inferior temporal cortex to prefrontal cortex via the uncinate fascicle in rhesus monkeys. *Experimental Brain Research*, *76*, 473–484.
- Vogel, E. K. & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, *37*, 190–203.
- Vogels, R. (1999). Categorization of complex visual images by rhesus monkeys. Part 1: behavioral study. *European Journal of Neuroscience*, *11*, 1223–1238.

- Wang, G., Tanifuji, M. & Tanaka, K. (1998). Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. *Neuroscience Research*, *32* (1), 33–46.
- Webster, M. J., Bachevalier, J. & Ungerleider, L. G. (1993). Subcortical connections of inferior temporal areas TE and TEO in macaque monkeys. *Journal of Comparative Neurology*, *335* (1), 73–91.
- Webster, M. J., Bachevalier, J. & Ungerleider, L. G. (1994). Connections of inferior temporal areas TEO and TE with parietal and frontal cortex in macaque monkeys. *Cerebral Cortex*, *4* (5), 470–483.
- Webster, M. J., Ungerleider, L. G. & Bachevalier, J. (1991). Connections of inferior temporal areas TE and TEO with medial temporal-lobe structures in infant and adult monkeys. *Journal of Neuroscience*, *11* (4), 1095–1116.
- Weisstein, N., Montalvo, F. & Ozog, G. (1972a). Differential adaptation to gratings blocked by cubes and gratings blocked by hexagons: A test of the neural symbolic activity hypothesis. *Psychonomic Science*, *27* (2), 89–92.
- Weisstein, N., Montalvo, F. & Ozog, G. (1972b). Visual masking by interrupting gratings: Orientation specificity, depth cues, and absence of short-duration spread. *Journal of the Optical Society of America*, *62*, 9.
- Wilcox, T. (1999). Object individuation: infants' use of shape, size, pattern, and color. *Cognition*, *72*, 125–166.
- Wilson, F. A. W. & Goldman-Rakic, P. S. (1994). Viewing preferences of Rhesus monkeys related to memory for complex pictures, colours and faces. *Behavioural Brain Research*, *60*, 79–89.
- Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.
- Yukie, M., Hikosaka, K. & Iwai, E. (1992). Organization of cortical visual projections to the dorsal and ventral parts of area TE of the inferotemporal cortex. *Society for Neuroscience Abstracts*, 132.8.
- Yukie, M., Takeuchi, H., Hasegawa, Y. & Iwai, E. (1990). Differential connectivity of inferotemporal area TE with the amygdala and the hippocampus in the monkey. In E. Iwai & M. Mishkin (eds.), *Vision, memory, and the temporal lobe*. New York: Elsevier, pp. 129–135.

- Zhong, Y.-M. & Rockland, K. S. (2003). Inferior parietal lobule projections to anterior inferotemporal cortex (area TE) in Macaque monkey. *Cerebral Cortex*, *13*, 527–540.

Acknowledgements

First of all, I would like to thank my supervisor Dr. Gregor Rainer, for his continuous support throughout the whole work. Without his advice and help, the thesis would not have been possible. He provided the necessary guidance throughout the project, but also gave me the freedom to explore on my own. I'm also indebted to Prof. Nikos Logothetis for the possibility to work in the excellent Neuroscience environment he created in Tübingen. The possibilities in the lab seem unlimited, and I could learn a lot throughout the last years. Furthermore, I'm most grateful to Prof. Hanns Ruder for his interest in my work, and his willingness to accept me as a graduate student in the Physics Department at the Eberhard Karls University in Tübingen.

A large number of people have helped me to accomplish this thesis. It is not possible to mention all of them, but some have to be acknowledged in more detail. Within the lab, Mark Augath, Barbara Dillenburger, Asif Ghazanfar, Kathrin Guhr, Kari Hoffman, Christoph Kayser, Han Lee, Alexander Maier, Jon Pauls, Michael Schmid, and Melanie Wilke have provided special support. They have helped me with practical issues, discussed my ideas, provided me with their software code, and have thoroughly proof-read the manuscript. I also want to express my appreciation to Claudia Hoehne and Julia Wellsow for reading parts of the thesis.

Without Anne Pauls and Kerstin Stockmeier, accomplishing the Ph.D. thesis would have been half the fun. The fact that we went through the usual ups and downs together helped enormously. I could count on them to provide moral support, discussions, and entertainment whenever needed.

Finally, I owe a big thank you to my family and my friends for their support throughout the last years.