Neuronal basis of perceptual binding processes in visual object coding

Dissertation

in partial fullfillment
of the requirements for the degree

Doctor of natural sciences
(Dr. rer. nat.)

submitted to the Faculty of Physics,
Philipps University Marburg

by

Alexander Gail

Marburg / Lahn
August 2002

To Mareike

Preface

Our conscious experience is strongly dominated by the visual sense. *'I see', said the blind man* is more than a macabre joke. It emphasizes the close relationship between visual encountering of our environment and cognitive processes, which becomes manifest in the intriguing frequency of perceptual terms to describe our thinking processes. We believe that *Thinking is seeing* and think that *Seeing is believing*. Can you *see* the point of the argument? To *see* can mean to understand ("I see the point of the argument"), to imagine ("I can see him as president"), to recognize ("I see the problem"), to think or consider ("Let me see, ..."). Our *eyes* are used as metaphor for having an opinion or making an estimation ("in the eyes of the law"), drawing the attention to something ("catch someone's eyes"), being aware of something ("with one's eyes open"), and so on. What for do I make this linguistic excursion? It should emphasize that probably vision research is a more general approach than just the interest in one special sensory modality. It may help to understand why so many brain researchers work in the field of visual perception. I am far from reducing the human to a 'seeing-machine'. But I think that vision research is a useful and import part of solving the brain puzzle. There is good reason to believe that our knowledge of the neuronal basis of vision will help to gain a more general insight into the conception of our brain.

The physiological findings of universal anatomical and morphological building blocks of the brain justify the hope that processing mechanisms in the brain may be of general validity. Certain types of neurons and their connectivity, or the cortical layer structure can be found across the cortex. Why should the brain use the same structures for different ways of processing? Basic tasks in vision, like integrating distinct elementary information (binding) into larger conceptual units (objects), are not unique to vision. Equivalent tasks have to be solved for perception in other sensory modalities or in higher cognitive tasks, like problem solving. But there are also different demands on neuronal processing in different sensory or motor modalities or for certain cognitive tasks. Defining universal tasks on the one hand and differentiating specific demands on the other hand is important to avoid over-generalization. I consider the prementioned binding task for building conceptual objects a universal task.

The close relation between vision and cognition certainly has been one of the reasons for the historical interest of psychology in vision. The intensive use of visual paradigms in psychophysical research now serves as a useful background for neurophysiological investigations. The laws of visual perceptual organization, as formulated by the Gestalt psychologists in the early 20th century, or many optical illusions, which demarcate the limits of our visual capabilities, still need to find their neurophysiological explanation. The variety of psychological findings on vision offer good starting points for building theories on visual processing and deriving hypotheses to be proved with physiological techniques. The connectionist approach, relying on the connectivity in neuronal structures as basis for processing operations, offers in this respect more synergetic potential than an abstract information-processing approach, analysing mental processes detached from their neuronal basis.

There are other, more practical reasons for the use of the visual system as a research item as well. Visual stimuli are easily applicable and well definable to make a system-theoretical approach in analysing brain functions. Considering the brain as an input (sensory stimulus) – output (conscious percept or action) system makes possible the use of analysis tools originally developed for the description of technical systems. These kinds of tasks can easily be defined for the visual sense: "Look at this stimulus and tell me what you see." Of course it is not *that* easy. But compared to the auditory, somatosensory, or olfactory sense, the ways of reporting on visual perceptions seem to be more elaborate.

Another advantage of the visual system is that there are animal models, like macaque monkeys, which share most capabilities of human vision and its anatomical/physiological substrates. They therefore allow investigation of the electrophysiological mechanisms of visual perception on the neuronal basis. This is especially attractive since about one third of the macaque brain is dedicated to visual processing (which underlines the dominance of the visual sense from a phylogenetic point of view), and many of the cortical areas involved in vision are comparably easily accessible for electrophysiological recordings.

The techniques for non-invasive neurophysiological investigations, like functional magnetic resonance imaging (fMRI), are rapidly improving. These methods can be used in human studies, and their improvement, together with other not yet known developments, may someday make animal studies redundant. Up to now animal research is indispensable. We still need to understand the relationship between the electrical activity of individual neurons or local groups of neurons, which is accessible with the widely used and far developed electro-

physiological methods, and the more indirect signals registered with non-invasive techniques, like for example the blood oxygenation level in fMRI. For some fundamental questions on neuronal processing disentangling of the local circuitry, i.e., the functional interaction among small groups of neurons, is essential, and hence microelectrode recordings are still necessary. For other questions, e.g., concerning higher cognitive processes, animal studies will be insufficient. Hence human studies will be essential, and methods that are ethically justifiable for this purpose, like fMRI, EEG, etc. After all it is not the monkey brain we are primarily interested in, but the human.

Marburg, August 2002

Alexander Gail

Acknowledgements

Zusammenfassung

Die Hypothese der Bindung-durch-Synchronisation zur neuronalen Kodierung von Objekten geht von der Annahme aus, dass die Neurone, die dasselbe visuelle Objekt repräsentieren, durch synchrone Aktivität gekennzeichnet sind, während die Neurone, die der Repräsentation verschiedener Objekte angehören, keine synchrone Aktivität aufweisen. Um diese Hypothese zu testen, wurden neuronale Gruppenaktivitäten gleichzeitig mit mehreren Mikroelektroden aus der primären Sehrinde (V1) wacher Rhesusaffen aufgezeichnet, während diese visuelle Aufgaben ausführten. Die Datenanalysen konzentrierten sich dabei auf zeitliche Signalkorrelationen in verschiedenen Frequenzbereichen gemessen anhand der spektralen Kohärenz auf Basis lokaler Gruppen-Impulswahrscheinlichkeitsdichten (multiple unit activity, MUA) und lokaler Feldpotentiale (LFP).

Die Gültigkeit der Bindung-durch-Synchronisation-Hypothese wurde in einem ersten Schritt anhand eines Figur-Hintergrund-Reizes überprüft, welcher aus einem Leuchtdichtekontrast-Gitter zusammengesetzt war, bei dem ein Objekt durch räumliche Phasenverschiebung eines quadratischen Teilbereichs hervorgehoben wurde. In Übereinstimmung mit der Hypothese spiegelte die Kohärenz der LFP im γ-Frequenzbereich (35–90 Hz) die perzeptuelle Szenensegmentierung wider: Die LFP-Kohärenz war hoch innerhalb von Segmenten, d.h. innerhalb des jeweiligen Repräsentationsbereichs von Objekt und Hintergrund, und niedrig zwischen Segmenten, d.h. über den Repräsentationsbereich des Objektumrisses hinweg. Dies galt für späte neuronale Signalabschnitte (Latenz > 100–150 ms), nicht jedoch für frühere Abschnitte, und auch nicht für die LFP-Kohärenz bei niedrigen bis mittleren Frequenzen (1–20 Hz) zu irgendeiner Zeit, und nicht für die MUA-Kohärenz bei irgendeiner Frequenz zu irgendeiner Zeit. Zeitliche Kreuzkorrelationsanalysen der MUA ergaben qualitativ ähnliche Ergebnisse wie die LFP-γ-Kohärenz, allerdings mit geringerer statistischer Sicherheit. Eine Zunahme der lokalen Impulsrate im Repräsentationsbereich des Objektes im Vergleich zum Hintergrund, wie sie von einer weiteren Hypothese zur Objektkodierung vorgeschlagen wird, wurde in einem, nicht aber dem anderen der beiden Affen gefunden.

In einer zweiten Experimentserie konnten wahrnehmungsbezogene Modulationen der neuronalen Synchronisation mit Hilfe der binokularen Rivalität nachgewiesen werden. Beobachter nahmen bei dieser Aufgabe abwechselnd einen von zwei widersprüchlichen Gitterreizen wahr, die mit jeweils orthogonalen Orientierungen jedem Auge einzeln, aber gleichzeitig, präsentiert wurden. Entgegen der Erwartung, traten bei dieser Art der Stimulation in V1 wahrnehmungsbezogene Modulationen der LFP-Leistungsdichte und -Kohärenz bei niedrigen und mittleren Frequenzen (4–12 und 12–28 Hz) auf, aber nicht bei γ-Frequenzen (28–90 Hz), und dies kurz ($\leq$ 100 ms) vor der Verhaltensreaktion der Affen. Entsprechende, wahrnehmungsbezogene Modulationen von Maßen beruhend auf MUA wurden nicht gefunden.

Der Nachweis entkoppelter LFP-γ-Aktivität zwischen Figur- und Hintergrund-Repräsentation in V1 (Experiment 1) deutet darauf hin, dass diese Signalanteile eventuell die Figur-Hintergrund-Trennung unterstützen. Die Ergebnisse in Bezug auf binokulare Rivalität (Experiment 2) erfüllen nicht die Erwartung, dass γ-Synchronisation in V1 mit dem gegenwärtigen, subjektiven, perzeptuell bewußten Empfinden zusammenhängen, sondern legen stattdessen eine wahrnehmungsbezogene Bedeutung der Synchronisation bei niedrigen und mittleren Frequenzen in solchen rivalisierenden Wahrnehmungssituationen nahe.

Das modifizierte Wheatstone-Stereoskop zur dichoptischen Stimulation, das für das Experiment zur binokularen Rivalität entwickelt wurde, wurde mit Hilfe von dichoptischen Mehrkanal-Messungen der klassischen rezeptiven Felder (DCRF: dichoptical classical receptive field mapping) im On-line-Betrieb kalibriert, um eine optimale Ausrichtung der Augenvergenz und eine Minimierung der Fehlanpassung zwischen Vergenz und Akkomodation in den Affenexperimenten zu erreichen. Die Inkongruenz zwischen binokularer und links/rechts monokularer Stimulation konnte dadurch im statistischen Mittel auf ~0.025° Sehwinkel reduziert werden. Die Effizienz der DCRF-Methode wird am Beispiel von Kurzzeit-Aufnahmen zur Bestimmung der okularen Dominanz in MUA- und LFP-Daten demonstriert. Darüber hinaus eignet sich die Messanordnung in Verbindung mit der DCRF-Methode zur Untersuchung der links-rechts-okularen Interaktion bei der Bildung binokularer rezeptiver Felder als neuronale Basis des Stereosehens.

Abstract

The binding-by-synchronization hypothesis for neuronal object coding proposes that those neurons which participate in the representation of the same visual object are characterized by synchronous activity, while those neurons which belong to the representations of different objects do not show synchronous activity. In order to test this hypothesis, neuronal group signals were recorded simultaneously with multiple microelectrodes from the primary visual cortex (V1) of awake rhesus monkeys performing visual tasks. Data analyses concentrated on temporal signal correlations in different frequency ranges by means of spectral coherence based on multiple unit activity (MUA) and local field potentials (LFP).

The validity of the binding-by-synchronization hypothesis first was addressed with a figure–ground stimulus composed of a luminance-contrast grating, where an object defined by a spatially phase-shifted square sub-region was distinguishable. In accordance with the hypothesis coherence of LFP in the γ-frequency range (35–90 Hz) reflected perceptual scene segmentation: LFP γ-coherence was high within segments, i.e., within the respective cortical ranges of representation of object and background, and low between segments, i.e., across the representation of the object's contour. This was true for late neuronal response epochs (> 100–150 ms latency), but not for earlier epochs, and also not for LFP coherence at low-to-medium frequencies (1–20 Hz) at any time, nor for MUA coherence at any frequency at any time. MUA temporal cross-correlation qualitatively revealed similar results as LFP γ-coherence, but less reliably. Multi unit spike rate enhancement at the object's compared to the background's representation, as proposed by a complementary hypothesis on object coding, was found in one monkey, but not the other.

In a second series of experiments perception-related modulations of neuronal synchronization could be shown in a binocular rivalry task, where observers perceived alternations among conflicting stationary grating stimuli presented singly, but simultaneously, to each eye at orthogonal orientations. Against expectation, with this kind of stimulation perception-related modulations in V1 occurred in LFP power and coherence at low and medium (4–12 and

12–28 Hz), but not at γ-frequencies (28–90 Hz), shortly (≤100 ms) before the monkeys' behavioural response. Corresponding perception-related modulations of MUA-based measures were not found.

The finding of decoupling of LFP γ-components among figure and ground representations in V1 (experiment 1) suggests that these signals may support spatial figure–ground segregation. The results with binocular rivalry (experiment 2) do not support the expectation that γ-synchronization in V1 is related to the current subjective perceptual state of awareness, but instead suggest a perception-related role of synchrony at low and medium frequencies in such rivalrous viewing situations.

The modified Wheatstone stereoscope for dichoptical stimulation, developed for the binocular rivalry task, was calibrated on-line by means of multi-channel dichoptical classical receptive field (DCRF) measurements in order to realize optimal eye vergence alignment and minimal vergence–accomodation mis-match in the monkey experiments. Thereby the remaining incongruence of binocular and left/right monocular stimulation could be reduced on average to ~0.025° visual angle. The effectiveness of the DCRF-technique is demonstrated by short-term recordings for measures of ocular dominance in MUA- and LFP-data. Additionally, the setup combined with the DCRF-method is suitable for investigation of left and right eye interactions in forming binocular receptive fields as the neuronal basis of stereovision.

Statement of Originality

Major parts of this thesis consist of manuscripts already published (Chapter 2) or submitted for publication (Chapters 3 and 4) written in co-authorship with Reinhard Eckhorn and Hans Jörg Brinksmeyer. Especially Chapter 2 profits from contributions of Hans Jörg Brinksmeyer to illustrations, and from textual drafts of the Introduction and Discussion sections by Reinhard Eckhorn. All other parts of this thesis are originally composed by the author unless otherwise indicated explicitly.

The experimental setup was largely established by the author, including the development of a completely re-newed software environment for visual stimulation with real-time interactive control of the monkey behaviour and automated recording. The author carried out the monkey training, in essential parts performed surgery under surveillance of Prof. Dr. Roman Bauer and Prof. Dr. Reinhard Eckhorn, and conducted the recording experiments providing the data contained in this work; surgery and experiments were jointly carried out, primarily with Hans Jörg Brinksmeyer. Data analyses were implemented self-dependently, including extensive software developments; elementary analysis tools were written in close co-operation with Andreas Bruns.

Contents

1 Introduction

How do we manage to break down the complexity of a visual scene into recognizable subunits without noticeable effort? Or the other way round: How do we integrate the tremendous amount of distributed local pictorial information into meaningful objects? These are basic and unsolved questions on the functionality of human vision. Both questions refer to the same problem of seeing but constitute different ways of seeing the problem. What is common to both questions is how our visual system extracts relevant information and reduces it to a comprehensible amount, and especially how this information is represented in our brain.

Perceptually a visual scene is composed of a variety of *visual objects*. What makes a visual object an object in terms of the electrical signals of the many distinct neurons involved in processing the visual information contained in the stimulus? How can the information constituting an object dynamically be integrated within a highly connected network of neurons to form a coherent percept of this object? This is the essence of the *binding problem*. From the psychophysical point of view the present work deals with the binding problem in visual object perception with regard to *figure–ground segregation* and *visual awareness*. From the perspective of neuronal coding the main focus lies on the role of dynamic assembly formation by *synchronization* of neuronal activity.

Before I will pinpoint the specific formulation of the working hypothesis on neuronal object coding being subject of the experimental studies presented here (Chapters 2 to 4), I will try to make clear why at all there is a binding problem that needs to be solved. It is therefore necessary to consider more closely the nature of visual objects and some basic properties of the neurophysiology of the visual system (cf. below).

1.1 Visual object perception

I will beware of trying to give a definition of what a *visual object* is. Most people have an idea of what it is, and this to a far extent is just what I mean when I use the term. A really sat-

isfying definition, from the scientific point of view, is most probably not possible at the current state of visual neuroscience. And it will not be possible until we have understood how object vision is achieved by our brain, or until we have succeeded in teaching computers to see with comparable performance. Nevertheless the next paragraphs should help to confine the concept of a visual object.

Visual objects. When we enter a room and are invited to describe what we see, we certainly will answer something like "a chair", "a table", "a bottle", and so on (Fig. 1.1A). We consciously experience our visual environment in terms of a composition of sub-units: visual objects. In the real world visual objects correspond to physical entities which typically have a meaning to us, and which we assume exist in the distal world, since they are ascertainable with different sensory modalities or may even be manipulable. We will answer in the same way in concrete terms when we look at a representational painting, although the depicted objects may or may not exist in reality, but still have a meaning to us and can be categorized. In the case of the impressionistic painting (Fig. 1.1B) the visual stimulus consists of blobs of paint on a surface. The arrangement of the blobs does not cause the same pattern of light on our retinas as a corresponding physical object would do, but the pattern resembles it, which is enough to remind us of such an object and to categorize it. We will mostly fail to give concrete answers like above when we look at an abstract painting (Fig. 1.1C). But still we tend to
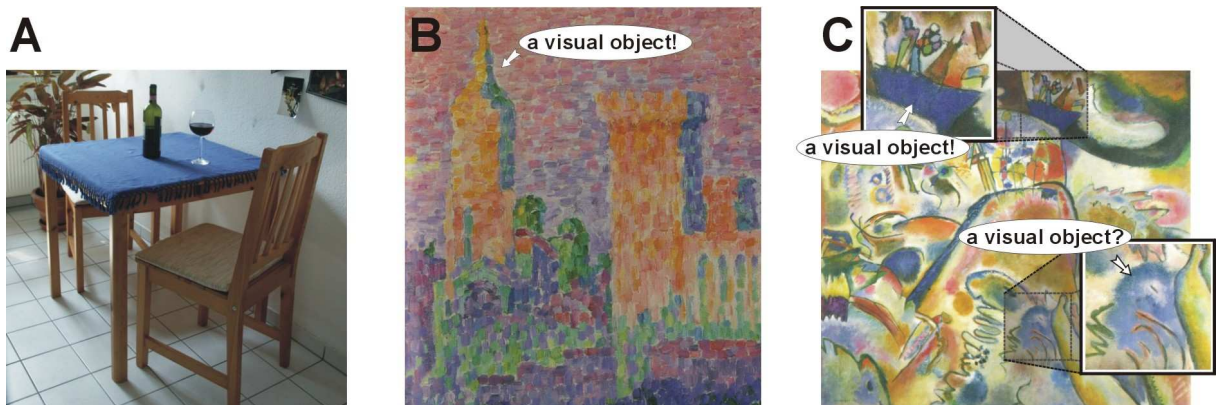
Figure 1.1 Phenomenological examples to delimit the meaning of the term *visual object*. **A:** In the real world visual objects as a rule refer to physical objects, which are ascertainable with several senses, may be manipulable, and mostly have a meaning to us. **B:** Depictions of realistic objects in representational paintings may completely differ from their real world counterparts in the pattern of visual sensations they cause. But still they usually appear to us as perceptual units and can be recognized – a fact *Pointilism* made extensive use of (Detail from Paul Signac, Le château des Papes à Avignon, 1900; reproduced from Feist (2000)) **C:** With increasing degree of abstraction, artists of the classical modernism approached the limits of visual object perception (Wassily Kandinsky, Kleine Freuden, 1913; reproduced from Düchting (1992)). There is no definite answer of what counts as visual object and what does not. A visual object may be understood as a perceptual entity based on visual sensations of which we can build a concept or mental representation without necessarily being able to name or categorize it.

describe the scene in terms of geometrical figures, like "a blue distorted rectangle", "a blue splash" etc., as far as possible. At this point the nature of visual objects becomes less obvious. There is no more relation to real world objects or predefined concepts, but still there is structure in the visual scene and we can extract isolated units and conceptualize them. I will refer to such perceptual units also as visual objects.

The previous paragraph has given examples, but what are more general characteristics of visual objects? In any case a visual object seems to occupy a limited partition of the visual scene; it can be localized which means that we can tell where it is and how big it is. A bird in front of the blue sky is easy to be localized, and likely to be seen as an object (Fig. 1.2A). The sky, on the other hand, is assumed to extend over almost the entire visual field, especially even behind the bird. This is an example of *figure–ground segregation*. While the figure is limited in space the background typically is perceived (although not seen) as spreading behind the object. What makes the potentiality of an object to become localized? For the bird, not for the sky, we can follow its contour, which encloses a single coherent area (Fig. 1.2). More formally, the potentiality to become localized may be described as a net convexity of the outlining contours resulting in a kind of *closure* (cf. also Fig. 1.3F). But the possibility of localization can not be sufficient for an object definition. For example, the role of the background is relative, as has been demonstrated with Escher's birds (Fig. 1.2B). Partial occlusion (e.g., a dog behind a lamp post) can visually divide an object into separate parts and thereby prevent unique localization in terms of a single coherent area. Also, distinct parts of an object
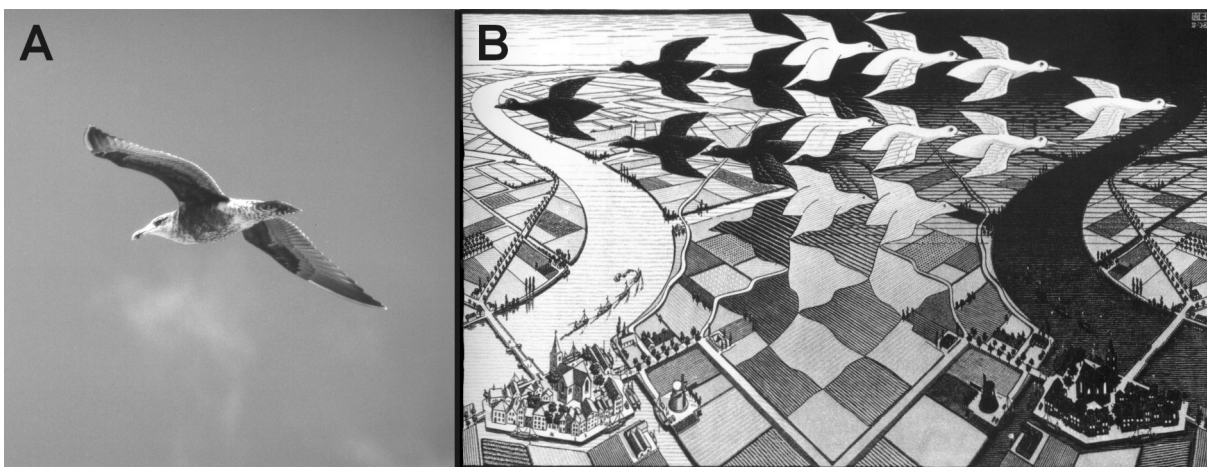
Figure 1.2 The potentiality of visual objects to become localized is a prerequisite for *figure–ground segregation*. **A:** In the case of a single bird in the sky, localization is an easy task for our visual system. **B** illustrates how the potentiality of an area to be localized contributes to perceiving this area as a visual object. As soon as the outline of an area is not closed anymore, while the outlines of adjacent areas are, the closed areas are likely to become objects, and the others to become background (M.C. Escher, Day and Night, 1939; reproduced from Ernst (1992)).

can themselves become localized in isolation (the beak of the bird, a wing of the bird, a feather of the wing). The same part of the scene can be an object by itself or part of an object, which means that objects can be hierarchically organized. Hence, visual objects can additionally be characterized by the grouping of parts of the scene into larger units (the two halves of the dog; the beak, the wing, etc. of the bird). This process is called *perceptual organization*. The outcome of this process not necessarily has to represent a known or meaningful object to be considered a visual object, as is the case for novel or abstract figures. Perceptual organization is not arbitrary but follows certain rules. These rules will help to further circumscribe the concept of a visual object.

Cognitive aspects of object perception. One way to learn more about the nature of visual objects, and hence the work style of our visual system, is to ask for the rules of perceptual organization. It seems that there is no complete set of rules that allows to clearly define what makes a visual object: Visual objects in general can not be defined by their pure visual sensory properties. This insight was furthered by the present incapability to build an artificial visual system with a performance comparable to human vision.

Already before the invention of computers (and the sobering realization of their apparent object agnosia) the Gestalt psychologists (Wertheimer, Koffka, Köhler; e.g., Koffka, 1935)

A *pregnanz*   **B** *good continuation*   **C** *proximity*   **D** *similarity*   **E** *common fate*

F *convexity*   **G** *symmetry*
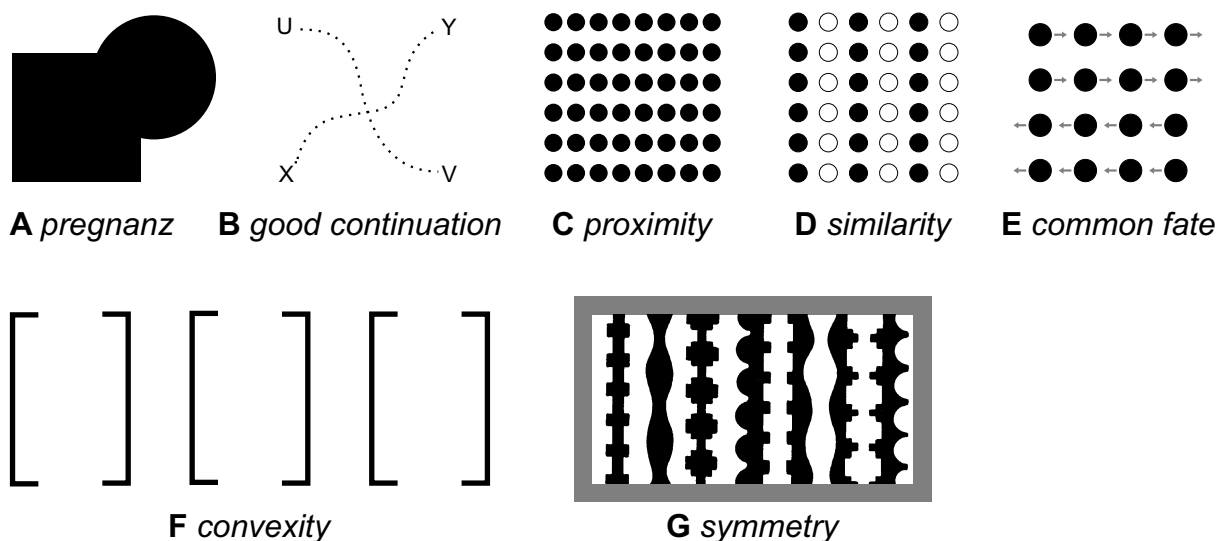
Figure 1.3 The Gestalt laws describe principles of *perceptual organization*. Especially the law of *pregnanz* (also: *good figure*; A) and *good continuation* (B) underline the interpretative character of object vision in the constructivistic Gestalt approach. For example, the preferred interpretation of the scene in A as being constituted of a square and a circle, although neither of these is physically present in the visual stimulus, reveals a cognitive contribution to the process of visual object perception. From visual experience the square and the circle are more familiar, and this knowledge biases interpretation. The preferred grouping due to *proximity, similarity, common fate* (movement in the same direction), *convexity* and *symmetry* (C–G) illustrate basic rules for figure–ground segregation and object integration.
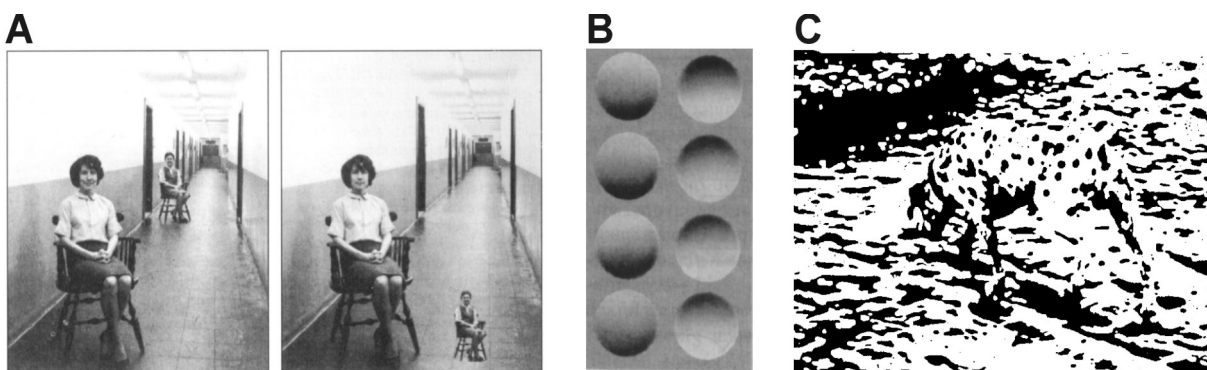
Figure 1.4 *Experience* and *knowledge* influence object perception. **A:** The perceived size of visual objects depends on the estimated distance, which is an unconscious conclusion from real world experience. **B:** The 3-D interpretation of the shaded spherical objects preferentially attributes heightening to the left column and deepening to the right. The inverse interpretation is harder to achieve although the objects are mirror-symmetrical. The reason for this bias is our internal assumption about illumination typically coming from above (the interpretation reverses when turning the picture by 180°, and vanishes at 90° rotation). In A and B the detection of the objects does not depend on cognitive influence, but the estimate on certain attributes does. **C:** In difficult viewing situations even detection (and hence object definition) may depend on previous knowledge. If you have never encountered this picture before it will be hard to perceive this famous actor of the visual neuroscience community (A, B: reproduced from Kandel et al. (1996); C: photograph by R.C. James; reproduced from Goldstein (1989)).

formulated systematic rules of perceptual organization. Their Gestalt principles (Fig. 1.3) are mainly founded on introspection and mark the transition from structuralism to constructivism. *The whole is more than the sum of its parts* was the basic principle. Perception was recognized to be an active process rather than a passive collection of sensations[1]. This means, to become a percept a sensory stimulus is subject to organization (according to rules) and, beyond this, to interpretation. The interpretative character of perception can be demonstrated by many optical illusions and by the knowledge-dependence not only of object recognition, but also of detection (Fig. 1.4). In conclusion, visual objects are the result of cognitive processing based on experience and previous knowledge about the world.

Even by usage of our visual experience and knowledge a sensory stimulus may not deterministically lead to a definite perception. This can be elucidated with ambiguous figures (Fig. 1.5): The interpretation of the scene is permanently changing between two possible solutions although the sensory information does not change. Perceptual organization, and therewith visual object definition, is a continuous and dynamic process.

Visual features. Although in most everyday situations objects come to the mind first, we are able to identify the building elements of a visual object. We can describe a bottle of wine as being green, having a smooth glossy surface, having parallel borders with a conic transition towards the bottle neck, etc. Visual objects are a composition of such basic *features*. The rela-

---

1 'Active' vision does not mean that we are aware of what is going on or have to guide processing. We are only aware of the outcome of this processing, i.e., our current perception, not of the way we achieve it.
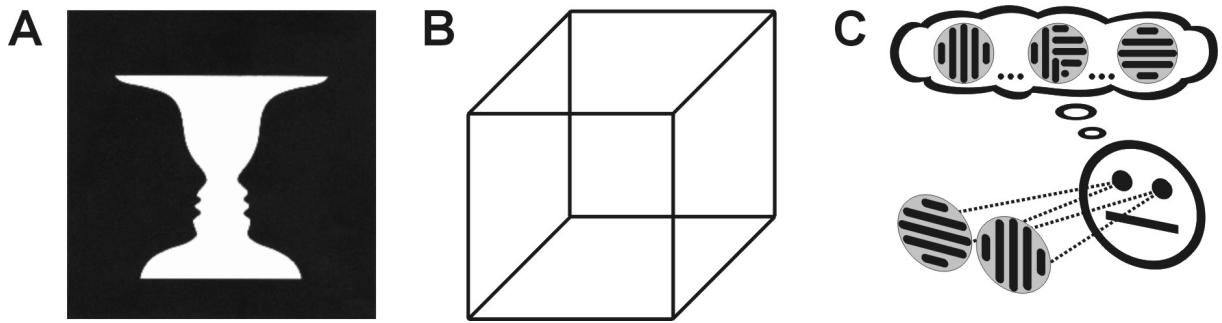
Figure 1.5 Visual object perception is a dynamic process and not deterministically dependent on the physical stimulus. When ambiguities in the interpretation of the visual scene can not be solved by experience or knowledge, perception vacillates between both solutions. **A:** *Rubin's Vase*. The figure–ground segregation is ambivalent in this example: Either a white vase in front of the black background can be seen, or the silhouettes of two faces on a white background. **B:** *Necker's Cube*. The depth interpretation of the skeleton cube is ambivalent: The front-side is either given by the upper right or the lower left square. **C:** *Binocular rivalry*. Presentation of incongruent conflicting stimuli to the two eye makes perception change between both stimuli irregularly over time. The experimental benefit of such stimuli is the possibility to dissociate between neural processes related to conscious perception (which alternates) and sensory stimulation (which does not change). Neuronal activities being modulated in correspondence with changes of the subjective perceptual state despite constant sensory stimulation are often referred to as *correlates of visual awareness*.

tion between the subjective *pop-out* (Ulric Neisser) of objects as a whole and feature identification was intensively studied by Bela Julesz (1986) and Anne Treisman (1996, 1998). Object perception can be seen as a process, where information based on distinct features (*perceptual primitives* according to Treisman) is bound together to form an overall impression. This process is called *feature binding*. Binding refers to several types of temporary integration processes, of which *spatial binding* is one prominent example (also referred to as *part binding*). Spatial binding is not restricted to the integration of discrete elements but includes integration of continuous surfaces. The result of spatial binding is *segmentation* of the visual scene into separate partitions. This is, for example, what happens when we segregate figure from background. The need for another sort of binding results from the fact that different attributes of an object often share the same location, as for example the colour and the texture of an object's surface do. This means that features of different quality have to become bound to constitute an object (*property binding;* overview: Treisman, 1996). In this view visual objects are essentially defined by the relations among different visual properties and their respective locations, i.e., visual objects are the outcome of various binding processes.

1.2 The binding problem

The idea of visual objects being the result of a process of feature integration is somewhat at odds with the subjective impression of objects being instantaneously present as a whole when we look at them. Consciously we experience the parsing of an object into more elementary units as a post-hoc analysis which requires attention. In contrast, we are not aware of an integration process. Why do we – from the scientific point of view – nonetheless think that visual objects become actively bound together from a composition of features, and why does there result a binding problem from this view? What neuronal mechanisms could be helpful to overcome the problem?

Functional specialization in the visual system. The key to the binding problem is the fact that different features of a visual object (like colour, shape, motion) are likely to be processed and represented by distinct parts of the brain, i.e., distinct neuronal units. There is converging anatomical-physiological, psychophysical, and neuropsychological evidence supportive for this notion of functional specialization. Even within each single feature dimension (e.g., line orientation) parameter values (e.g., horizontal or vertical) are represented by the activity of distinct (groups of) neurons within and across several stages of processing (cf. below). A visual object typically consists of different features and contains more than one value of a given feature. For example, it covers an extended spatial range with a changing surface brightness due to shading, and is limited by contours of varying local orientation. All the distributed information needs to be integrated correctly to allow coherent object perception and reliable identification. The problem is aggravated by the difficulty to assign the right colour, surface texture, and contour to the respective objects in a typical everyday scene where multiple objects overlap.

Before I will refer to proposed neuronal mechanisms of integration potentially suitable to solve parts of the binding problem, I will outline some evidence on separated processing of visual features.

Anatomical-physiological evidence of separated feature processing. The visual system comprises sub-cortical structures and many cortical areas. In monkey visual cortex more than 30 visual areas are distinguished (Fig. 1.6). Different areas are commonly believed to be especially sensitive to different visual properties, like motion (area MT, MST) or colour (V4). Visual processing in the cortex has functionally been divided into separated pathways, a *What-*

and *Where-path* (Ungerleider and Mishkin, 1982), each including several visual areas. The *dorsal* (*parietal*) Where-path is meant to be mainly engaged in localising visual objects, directing visual attention, and steering visuo-spatial behaviour, while the *ventral* (*infero-temporal*) What-path seems to be more in charge with object identification. These functional classifications are not to be taken too strictly, yet denote an asymmetry in stimulus specificity and task responsibility, and hence indicate separated processing of different object properties to a certain degree.

Visual areas are not isolated processing modules, instead they constitute a complex network (Fig. 1.6). The architecture of anatomical connections between visual areas suggests a co-existence of hierarchical and parallel processing with massive options for interactions. The connections between areas are mostly reciprocal, allowing *feedback interaction* along the hierarchical forward (*bottom-up*) – backward (*top-down*) axis, and between parallel process-ing streams (*lateral*). Within each area the representation of visual features builds special to-pographies based on parallel processing combined with feedback and lateral interaction. The latter can be mediated either directly by a dense network of local (intra-areal) lateral connec-tions (e.g., Gilbert, 1993), or indirectly via intra- or inter-areal feedback loops containing con-vergent or divergent projections (review: Salin and Bullier, 1995).

For examining the neurophysiological background of the binding problem, I want at this point to focus on separated processing of visual object properties. This view has primarily been gained by demonstrating selective responsiveness of neurons to different visual stimuli pre-sented in isolation. The selectivity with respect to visual features (like location, colour, etc.) finds its expression in the concept of the *classical receptive field* (**CRF**).

A dense layer of *photoreceptors* ($\sim 10^8$) samples the visual field ($180° \times 130°$ in human binocular vision) geometrically projected onto the retinae by the lenses. The parallel process-ing with respect to spatial information is kept along the lower (in the sense of 'more periph-eral') main visual pathway via *retinal ganglion cells*, *optic nerve, lateral geniculate nucleus* (**LGN**), up to the first cortical area, the *primary visual cortex* (**V1**; also: *striate cortex, Brod-man area 17*) and some other areas. This organization principle of *retinotopic* projections (keeping the retinal topography except for spatial distortions) finds its expression in spatially confined CRFs: Visual neurons in retinotopically organized areas are directly affected in their activity only by light changes in a restricted partition of the visual field. Retinotopy constitutes a continuous but distributed neuronal representation of the basic feature *visual field location*.

Retinal processing is not only parallel with respect to space but also to colour selectivity. Two functional kinds of photoreceptors, *cones* being colour selective, and *rods* showing broad spectral sensitivity, represent an early division in the visual system. The separated signalling of colour is also maintained over several stages. Already at the level of retinal ganglion cells it is combined with another functional distinction: the *magno-* and *parvo-cellular pathway*. The parvo-pathway is characterized by high spatial resolution, low temporal resolution, and colour selectivity; the magno-pathway shows no colour selectivity, lower spatial resolution, but higher contrast sensitivity, and higher temporal resolution. Each pathway contains small and large retinal ganglion cells, which themselves receive input by both rods and cones via *bipolar cells* (laterally connected by *amacrine cells*). The two pathways are named after the clearly

Figure 1.6 Anatomically defined sub-cortical stages and cortical areas, and their connectivity in the monkey visual system (Rees et al., 2002; modified from: Felleman and van Essen, 1991). The layout of cortico-cortical connectivity among separate visual areas demonstrates hierarchical processing in partly parallel streams with extensive lateral and top-down feedback interactivity. The different cortical areas are associated with various forms of functional specialization.

separated magno- and parvo-cellular layers in the LGN they project to. The magno- and parvo-pathway partially interact at the level of V1. They project to different V1 input layers (4Cα, 4Cβ) but in part become recombined in the *blob regions* (stainable by cytochrome oxidase) of the upper V1 layers 2/3 (Livingstone and Hubel, 1984, 1988).

V1 is the locus of unification for another kind of early separation: Left and right eye *monocular* afferent inputs (ascending from receptor to central nervous system) converge at the level of V1 for the first time onto neurons with *binocular* receptive fields, which are excitable by visual stimulation via each eye.

As a consequence of separation on the one hand, and recombination and lateral interaction on the other hand, neurons in V1 can be selective with respect to many aspects of the visual sensory input: location, ocular origin (left or right eye), local orientation of luminance based contrast-borders and their direction of movement (e.g., Hubel and Wiesel, 1977), spatial frequency (e.g., DeValois et al., 1982), colour (Dow and Gouras, 1973), and stereoscopic

depth (retinal disparity between eyes; e.g., Barlow et al., 1967). Selectivity for the one or the other visual feature is differently pronounced in individual V1 neurons, but in general can be quite broad for features except location. This means that the selectivity for combinations of features with different qualities is weak in V1. Together with the spatially confined sizes of CRFs, which are small compared to the typical size of visual objects and strongly overlap in neighbouring neurons, one can conclude that many neurons in V1 are simultaneously activated by a single visual object, while even more other neurons are at the same time activated by the rest of the visual scene. This implies the necessity for defining subpopulations to distinguish between distinct visual objects.

The functional specialization of visual areas and the definition of separated processing pathways, as mentioned at the beginning of this paragraph, indicate that feature selectivity is not only a property of single neurons at early processing stages, but can partly be found as preferred sensitivity to the one or the other feature in the majority of neurons of an area at other stages of processing as well. (For more details on the functional structure of the visual system see, for example, Nicholls et al., 1992; Kandel et al., 1996.)

The feature selective response of cortical neurons as outlined in this paragraph suggest the necessity to integrate the neuronal representation of a visual object (binding) for perception, since the information constituting the object is distributed within a network of many separate neurons. Note that the isolated presentation of single small test stimuli used to derive the classical receptive field properties reduces contextual influences on the neuronal response to a minimum, and is likely to reveal mainly the spatially linear filter properties of neurons. The use of stimulus arrangements larger than the size of the CRF of a neuron in V1, for example, can reveal response characteristics of this neuron not covered by the CRF concept (e.g., Knierim and van Essen, 1992; Kapadia et al., 1995; Lamme, 1995; Zipser et al., 1996; Kastner et al., 1997; Zhou et al., 2000). Such *contextual modulations* of the neuronal response strength at the level of V1 have to be taken into consideration for understanding visual object coding, together with receptive field properties in visual areas beyond V1, and together with coding schemes not purely based on response strength but regarding temporal response characteristics of the neurons as well (cf. Section 1.3 below).

Psychophysical evidence of separated feature processing. Functional specialization in the visual system can also be demonstrated psychophysically. One hint on selective processing within a single feature dimension is *selective adaptation*. Viewing of a grating stimulus with a certain orientation (e.g., vertical; for about a minute) decreases contrast sensitivity to a succes-

sive presentation of the grating stimulus especially for the previous orientation (vertical), and less for other orientations. This is the psychophysical counterpart to the fact that differently oriented contrast borders are represented by distinct adaptable neuronal resources, i.e., neurons with receptive fields of different preferred orientations (cf. above). Other psychophysical experiments demonstrate selective processing between different feature dimensions, namely the partial independence of form, colour, movement, and depth perception (Livingstone and Hubel, 1987). Although motion perception, for example, is not 'color-blind', as early research suggested, stimuli defined by pure chromatic contrasts (*equiluminance*) in some situations are inadequate to induce reliable perception of fast motion. These findings indicate a qualitative difference of color and motion processing (review: Gegenfurtner and Hawken, 1996). Similarly was the independence of colour and depth perception suggested by an impairment to induce depth perception with stereoscopic depth or perspective cues formed by contrasts of equiluminant colours, while luminance contrasts serve well for this purpose (Livingstone and Hubel, 1987).

There are not only psychophysical indications for selective processing but also for the necessity to flexibly bind features of different quality for an adequate interpretation of a visual scene. In normal vision we usually do not suffer a problem in correctly assigning different characteristics to the corresponding objects, even when there are several objects adjacently present. We know what colour belongs to the shirt, and what colour to the trousers, when we look at a person next to us. But with very brief presentations misperceptions casually happen. Asked for the colours of a briefly flashed green X (the trousers) and a red O (the shirt) subjects sometimes report a red X and a green O (Treisman and Schmidt, 1982). In a series of experiments Treisman described the conditions under which such *illusory conjunctions* occur. The interpretation of the data is that without the help of *selective attention* (confined to a restricted *spatial focus*), which is prevented by the shortness of the presentations, we are not able to bind two qualitatively different features (here colour and shape) to the location (of shirt or trousers, respectively) they have in common. But this binding to location would be necessary for a correct perception since shirt and trousers are identified by means of their shape, which is processed independently from their colour. The location as a special feature serves as the connecting link in this view (**FIT**: *feature integration theory*; Treisman and Gelade, 1980; review: Treisman, 1998). Another line of evidence comes from a series of visual search tasks (reviews: Treisman, 1996; Wolfe and Cave, 1999). The results (Fig. 1.7) indicate that there is a major difference in processing spatially distributed information within one

feature dimension (line orientation) and across different dimensions (orientation and colour): With one dimension the search task is assumed to be solved for all potential locations at a time (*parallel search*), while the conjunction of two features can only be evaluated for one location after the other (*serial search*). *Conjunction search*, according to FIT, needs the help of selective attention directed to each location successively to allow binding the features of different quality at this location (Treisman and Sato, 1990).

Note that these topics, illusory conjunctions, conjunction search, and the role of attention for binding, are subject to long-lasting controversies (including model suggestions different from FIT, e.g., Wolfe et al., 1989; Ashby et al., 1996), which I will not recapitulate at this point. For the purpose of this introduction it is sufficient to realize that a) there is evidence on the psychophysical relevance of the physiologically defined separated processing pathways and b) there are widely accepted psychophysical indications for a binding problem.

Neuropsychological evidence of separated feature processing. Among the earliest hints on distinct processing streams for different visual features like colour or shape came from neuropsychological findings. Cerebral or cortical lesions of limited extent can cause a variety of visual perceptual disorders. *Visual agnosia* (Freud) refers to an inability to visually recognize objects or attribute their features although there is neither a primary sensory loss nor mental deterioration. Patients with *colour agnosia*, for example, can distinguish and name different colours, but can not tell to which object a colour belongs. Similarly there are patients who suffer *motion agnosia* and are unable to perceive the movement of objects. *Prosopagnosia* is a rare but curious deficit in recognising faces, while having no trouble with other visual objects. Different types of agnosia underlined the existence of cortical areas engaged with different
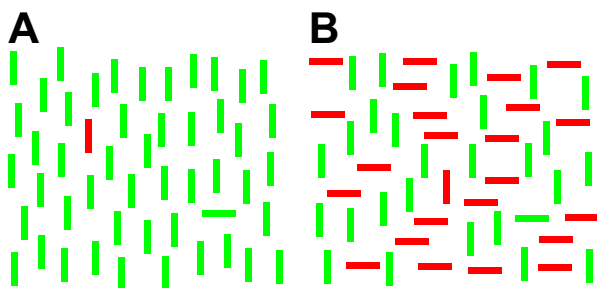
Figure 1.7 Psychophysical experiments can demonstrate how the necessity for binding features of different quality into an object influences our performance in *visual search* tasks (adapted from Wolfe and Cave (1999)). In this class of experiments subjects have to find a target (the red vertical or the green horizontal bar) within a field of *distractors* (e.g., green vertical bars in A). The time needed to solve this task strongly depends on the way the target can be distinguished from the distractors. **A:** If the difference is based on a single basic feature (colour in the case of the red vertical bar or line orientation for the green horizontal bar), search time is nearly independent from the number of distractors (*parallel search*). **B:** If the difference is based on the combination of at least two features (*conjunction search*; the red vertical or the green horizontal bar among red horizontal and green vertical bars), search time typically increases linearly with the number of distractors (*serial search*). In this case features of different quality (colour and orientation) have to be bound for each object to be distinguishable. This, according to *feature integration theory* (Treisman and Gelade, 1980), requires a serial scan of each object location by means of spatially *selective attention* to allow feature binding at the respective location.

visual tasks already before detailed neurophysiological data were available. Visual field deficits with a spatially restricted loss of sensitivity (*scotoma*) are indicative for the existence of a topographic cortical organization, i.e., a spatially ordered representation of the visual field. In contrast to agnosia these forms of blindness (*anopia*) can much easier be related to the anatomical-physiological organization of the visual system. A general difficulty in relating such deficits to cortical organization principles or neuronal mechanisms arises from the fact that lesions are often multi-focal with regard to anatomical-histological area definitions or hard to assign to them, and the numbers of well investigated patients are low. But the selectiveness of neuropsychological deficits elucidates the relevance of the anatomical-physiological findings for human perception and behaviour and confirms the existence of separated processing for different aspects of visual object properties. (For a more comprehensive overview of neuropsychological disorders see, e.g., Kolb and Whishaw, 1993).

1.3 Concepts on the solution of the binding problem

We have not only seen that there is a binding problem for perceptually integrating the features constituting an object, but we can also be sure that there is a solution for it. Why? Because in normal vision we do not suffer a problem in fast and robustly recognizing objects. Hence the search for neuronal mechanisms capable to overcome the binding problem is not an ill-defined problem. But it may be misleading to expect a single solution for the various demands on binding mechanisms. Scene segmentation, i.e., the spatial binding and segregation of different locations continuously represented in topographic cortical maps, may be neuronally implemented in a different way than binding colour to motion, two features processed by distinct visual sub-systems.

In the following I will describe different concepts proposed to solve the binding problem. These concepts should be seen as complementary rather than being considered mutually exclusive. *Convergent projections* ensure fast and robust integration of separated information, but lack flexibility if implemented in a hard-wired network. To increase the flexibility of a convergent wiring mechanism for object integration, it has to be combined with mechanisms of adaptively selecting partial information for further processing and flexibly combining separated information within a network. Such *selective processing* and *flexible binding* is needed

to account for the large number of feature combinations our visual sense is potentially confronted with, and to allow for variability of perceptual organization as occurring, for example, in ambiguous viewing situations (Fig. 1.5).

Based on neurophysiological findings different mechanisms supportive for unambiguous and flexible object definition have been proposed relying either on modulations of the neuronal response strength (i.e., average spike rate within ~50–500 ms), or on modulations of the temporal structure of discharge patterns in groups of neurons leading to fast synchronization (with a temporal precision of ~1–10 ms) of neuronal activities. Rate enhancement as well as increased coincidence of multiple inputs (as a consequence of fast synchronization) are suitable to increase the response probability of potential target neurons, due to their threshold dependent spike encoding with rather short integration times (~10 ms and lower) in the activated cortex (e.g., König et al., 1996). Hence both mechanisms are compatible with a scheme of selective convergent processing, but probably work on other temporal and maybe also spatial scales.

Convergent projections. A widely used mechanism for integration in sensory systems is based on convergent projections, i.e., multiple feed-forward connections projecting to the same neuron. If this target neuron builds the sum of all its inputs it will respond the stronger the more of and the stronger its excitatory input neurons are active. Therefore the target neuron can represent a special combination of features represented by its input neurons (*combination-coding cells*). This principle can be found along hierarchically organized visual areas, and finds its expression in increasingly complex receptive fields, especially in the ventral stream: In the spatial domain there is essentially concentric band-pass like coding of location in the retina and LGN, of contrast-border orientation as a prominent feature in V1 (e.g., Hubel and Wiesel, 1977), up to a specificity for complex constellations of contours, surface colours and textures in the infero-temporal cortex (like faces; Perrett et al., 1982). If such a hard-wired filter concept of hierarchical neuronal coding could be realized for each location and for an arbitrary number of potential feature constellations then there would be no binding problem. Each conceivable stimulus anywhere in the visual field could be represented by a few neurons with a localized and highly sophisticated stimulus specificity (*cardinal neurons*; Barlow, 1972). Such a coding scheme is extremely expensive with respect to the number of neurons and connections needed, and hence yields a problem of combinatorial exploding. Besides, there is no indication for a central area where any kind of distinct information of an arbitrary complex stimulus is represented. Instead there are highly interconnected, but also partly

separated, processing streams that seem not to anatomically converge onto a single site (Fig. 1.6).

One way the visual system avoids expensive coding of too large a number of potential feature combinations is building partly *invariant* representations. Receptive fields typically show a trade-off between spatial selectivity (prominent at 'low-level' processing stages) and pattern complexity (at 'high-level' processing stages): Receptive fields along the ventral pathway have an increasing spatial range of sensitivity (size of classical receptive fields; e.g., Maunsell and Newsome, 1987), but are also increasingly selective for specific complex feature combinations (e.g., Kobatake and Tanaka, 1994; Tanaka, 1996; Hegdé and van Essen, 2000). This means, neurons with receptive fields selective for complex stimuli characteristically are invariant, for example, to the exact location or size of this specific stimulus. But with respect to object definition this invariance-coding produces a binding problem: Within the large area covered by a high-level classical receptive field multiple visual objects of the scene can be contained. But then the preferred combination of features, for which the neuron is selective, does not necessarily belong to a single object, which is in contradiction to an unambiguous object representation and perceptually may give room to illusory conjunctions (cf. above).

Another problem inherent to any fixed convergent wiring scheme is the lack of flexibility in coding objects, which would be necessary to explain some cognitive aspects of vision, like the previously mentioned dynamic perceptual switching in ambiguous viewing situations (Fig. 1.5).

Spike-rate modulations for object definition. Mechanisms based on spike-rate modulations and proposed to reduce the binding problem are discussed, for example, in the context of selective attention. To overcome ambiguities resulting from invariance-coding visual selective attention was proposed to allow a space-based selection of single objects. The role of attention in this view is to support the representation of the currently attended object by means of spike rate enhancements through external influence from other cortical areas. This support gains the attended object an advantage in the competition with simultaneously present objects, the representations of which are suppressed by mutual inhibition which is continuously in effect between all current object representations (*biased competition*, Desimone and Duncan, 1995; Luck et al., 1997; Chelazzi et al., 1998; Reynolds et al., 1999; with respect to spatial selection this resembles the ideas of FIT, cf. Section 1.2). In effect such a spatially selective mechanism can be seen as practically shrinking the receptive field size to the extent of the attended object,

increasing the spatial selectivity in the ventral areas and hence reducing the binding problem (reviews: Desimone, 1998; Reynolds and Desimone, 1999; Wolfe and Cave, 1999; For an alternative approach to unambiguous object definition by externally guided dynamic routing of signals see, for example, Olshausen et al., 1993, 1995). It is not clear, though, how such a selection process could work on an input containing multiple adjacent or even overlapping objects without having already solved the segmentation problem, i.e., the binding problem to define potential areas (objects) for selection. Mechanisms of selective processing may be helpful to avoid illusory conjunctions when the segmentation of the scene is already achieved.

It is noteworthy at this point that the assumption of spike rates of single neurons in low-level areas (e.g. V1), being capable of representing simple localized features only, is misleading, if taken too strictly. Several forms of contextual spike rate modulations from outside the classical receptive field in V1 have been reported (Knierim and van Essen, 1992; Kapadia et al., 1995; Lamme, 1995; Zipser et al., 1996; Kastner et al., 1997), likely to be related to modulations of perceptual thresholds (stimulus detectability) in corresponding viewing situations (e.g., Field et al., 1993; Kovács and Julesz, 1993; Polat and Sagi, 1993, 1994). The contextual spike rate modulations includes neurons that can indicate the side (relative to their CRF position) of a luminance-defined object of either contrast, although the object defining cues are far outside the CRF (*border-ownership cells*; Zhou et al., 2000). How such complex response characteristics are achieved at early visual stages is not yet known. It is likely that facilitating and inhibitory interactions within a lateral network of orientation-specific connectivity are supportive for contour-coding (e.g., iso-orientation facilitation for co-linear contour elements and inhibition for parallel or cross-oriented ones; reviews: Gilbert, 1993; Gilbert et al., 1996, Kapadia et al., 2000). But the so far described degree of specificity of lateral connections is not sufficient to explain the complexity revealed, for example, in border-ownership coding. Additionally, feedback influence was proposed to facilitate activity particularly within the region covered by an object's surface- compared to background-representation at the level of early visual areas including V1 (Lamme et al., 1998; Hupé et al., 1998; Lamme and Roelfsema, 2000; Roelfsema et al., 2002).

A general problem with contextual spike rate modulations for object definition in early visual areas emerges from the fact that the response strength of a neuron at the same time strongly depends on the degree to which its classical receptive field properties are matched by the part of the current object falling into the CRF range. This means that spike rate alone is

ambiguous to represent both local feature identity and affiliation of this particular feature to an object. Contextual modulations can not solve the segmentation problem.

Synchronized assemblies for object definition. A different approach to solve the binding problem starts with the concept of neuronal *assemblies* proposed by Hebb (1949). Assemblies, in this view, are transiently formed groups of neurons that constitute a distributed representation of a current perceptual entity (e.g., a visual object). They allow flexible binding of arbitrary feature constellations. A neuron being part of one assembly and representing a certain feature in one situation may be part of another assembly in another situation, but still representing the same feature. The variety of combinations among elementary features in the visual world is accompanied by an equivalent potential of combinations in temporary neuronal assembly formation with this coding scheme. But with this kind of distributed representation a superposition problem arises. Which neuron belongs to which of the current assemblies? One needs an assembly-defining tag of the individual participating neurons for associating information within and dissociating between assemblies to enable unique representations of multiple objects. The *synchrony* of the activity among participating neurons was proposed to represent such an assembly-defining tag (Milner, 1974; Reitböck, 1983; von der Malsburg und Schneider, 1986), while the individual response strengths still represent the degree of matching between the stimulus and the filter properties of the neurons, i.e., their receptive fields. The *binding-by-synchronization* hypothesis states that distinct parts of the brain (distinct neurons) that represent the different characteristics of a perceptual unit (the features of a visual object) are bound together to form a transient neuronal assembly by means of their electrical activity being synchronized. According to the hypothesis this synchronized assembly represents the current perceptual unit (the visual object) as a whole, and thereby contributes to the solution of the binding problem. Based on experimental findings of temporally correlated neuronal signals in accordance with perceptual laws of grouping, synchronization as an indication of feature binding is preferentially discussed for activity in the *γ-frequency* range (30–90 Hz; Eckhorn et al., 1988; Gray et al., 1989; Engel et al., 1991a,b; Kreiter and Singer, 1996; Frien and Eckhorn, 2000; reviews: Eckhorn, 1999, 2001; Gray, 1999; Singer, 1999; Tallon-Baudry and Bertrand, 1999). But temporal correlations at other frequencies are also discussed in terms of co-operative processing and interactivity (e.g., Schanze and Eckhorn, 1997; von Stein and Sarnthein, 2000; Tallon-Baudry, 2001; Bruns and Eckhorn, 2002), and may be interpreted in

relation to other forms of binding, like *hierarchical binding* (Treisman, 1996) across processing stages.

The binding-by-synchronization hypothesis suggests a measurable indicator for what is bound in a given situation. One important step in proving this idea is to settle the perceptual conditions which do or do not lead to neuronal synchronization; or more specifically, to look for the occurrence of neuronal synchronization in concordance with perceptual binding. This is the purpose of my doctoral thesis. The focus here lies on the representation of spatial feature relations (*feature linking*) within temporally correlated assemblies of confined range (*association fields*) extending the concept of classical receptive fields (reviews: Eckhorn, 1999, 2001).

The binding-by-synchronization hypothesis per se does not solve the question of how the selection is achieved of what parts of a current visual scene should become bound. Beyond demonstrating the occurrence of neuronal synchronization in accordance with perception it is substantial in parallel to investigate the underlying mechanisms. This could help to put down the rules of perceptual organization to some corresponding rules for neuronal synchronization. Up to now this can hardly be done by means of electrophysiological recordings alone. Computer-based neuronal network simulations taking into account biological constraints are essential to solve this part of the puzzle. Such model approaches (Eckhorn et al., 1990; Eckhorn 1999, 2001) accompany part of this work in close co-operation (Saam et al., 2000; Al-Shaikhly, 2001).

1.4 Visual object perception – summary

Visual object perception consists of several distinct tasks. The features composing the visual scene must be detected and identified (feature detection and identification). Those features constituting a visual object must be selected and bound (segmentation and feature binding). Selection takes place according to certain rules of perceptual organization but also depends on cognitive interpretation. The visual object has to be recognized (object recognition) and must be made accessible to consciousness (visual awareness). These latter two tasks may not be seen as sequential; both are prerequisites for perception not to become a blind alley but to allow guidance of behaviour beyond reflex action. The filter properties of cortical neurons (receptive fields) seem suitable to represent features of different complexity in functionally spe-

cialized parts of the visual system, and hence may be responsible for feature detection and identification. Cortical mechanisms of feature selection and binding for object integration are still unclear. Feedback and lateral connections are likely to support mechanisms of selection and binding. Synchronization is a proposed indicator for perceptual binding being subject to investigation in the present work. Neuronal correlates of visual object awareness and recognition are still a mystery at large, certain aspects of which can be addressed empirically though. The present work will also deal with the awareness-relevance of neuronal synchronization.

1.5 The experiments – guide for the reader

The two main experiments of my study concentrate on different tasks of object perception, yet both investigate *spatial feature integration.*

The *first experiment* (**Chapter 2**) deals with the role of *synchronization in visual scene segmentation*: How are figural objects separated from the background of a scene or from other objects, and how are these distinct visual segments and their contours represented in the cortex? Does neuronal synchrony occur in accordance with the perceptual segmentation in this case? This question is addressed with a *figure–ground stimulus* without behavioural relevance for the monkeys, i.e., the stimulus is presented while the monkeys perform a fixation task.

The *second experiment* (**Chapter 4**) deals with the role of *synchronization for visual awareness*: In how far are those aspects of neuronal synchrony, attributed to coding an object, relevant for the conscious perception of this particular object? This question is addressed with a *binocular rivalry* paradigm: Presentation of ambiguous information to the two eyes leads to perceptual switching between the two possible interpretations of the visual information, although the physical stimulus itself does not change. The monkeys have to report on their current perception. The stimulus-independent perceptual switching allows a dissociation of stimulus- versus perception-related modulations of the corresponding neuronal activity.

To apply the binocular rivalry paradigm to awake monkeys special experimental hurdles had to be taken, one of which is precise *dichoptical stimulation.* The setup and techniques developed for this purpose, and helpful to meet further research demands, are introduced in **Chapter 3**.

Both experiments investigate the binding-by-synchronization hypothesis in the retinotopically organized primary visual cortex (V1) of rhesus monkeys. The figure–ground experi-

ment focuses on stimulus-dependent spatial feature binding and its neuronal correlates. The binocular rivalry experiment focuses on perception-dependent spatial feature binding and its neuronal correlates.

2 Figure–ground Segregation

Contour decouples γ-activity across texture representation in monkey striate cortex

Previous work on figure–ground coding in monkey V1 revealed enhanced spike rates within an object's surface representation, synchronization of gamma oscillations (γ: 35–90 Hz) in object and background regions, but no decrease in signal correlation across the representation of a contour. The latter observation seems to contradict previous statements on the role of γ-synchronization for scene segmentation. We re-examine these findings by analysing different coupling measures and frequency ranges of population activities potentially contributing to figure–ground segregation. Multiple unit activity (MUA) and local field potentials (LFPs) were recorded by parallel μ-electrodes in monkey V1 during stimulation by a grating in which an object was defined by a shifted rectangle. In contradiction to the conclusions in previous work, we find strong decoupling of population activity between figure and ground representations compared to the situation in which the object is absent. In particular, coherence of late γ-LFPs is strongly reduced, while reduction is absent during the early epochs of high amplitude transients for LFP- and MUA-coherence at all frequencies, and at low frequencies also in the subsequent epochs. Our results of decoupling in late LFP γ-components among figure and ground representations suggest that these signals may support figure–ground segregation.

The content of this chapter has been published in Gail et al., 2000. Preliminary versions had been published as conference proceedings (Gail et al., 1999a,b).

2.1 Introduction

Perceptual grouping and scene segmentation are basic aspects of early visual processing as has already been shown by Gestalt psychologists (Wertheimer, 1923; Koffka, 1935). These aspects are tightly connected to figure–ground segregation, a necessary requirement for object

recognition. Segregation is a crucial and difficult task because the same object can appear on a variety of backgrounds, including the most difficult situation, which is composed of largely similar features for figure and ground. While many psychophysical investigations provide material about perceptual feature grouping in various visual situations (reviews in: Boucart, 1999), neural mechanisms of figure–ground segregation are largely unknown.

Spike rate modulations for figure–ground segregation. A classical approach to figure–ground segregation assumes convergent forward connections to a common target of those neurons representing the specific features of an object by their receptive fields. A small number of single cardinal neurons (Barlow, 1972) or a subgroup of neurons of an associative memory (e.g., Fukushima, 1980; Palm, 1982) may represent the presence of an object by an increase in their activation. Model investigations suggest an object specific increase in spike rate which in turn might suppress activities evoked by other objects via mutual or common feedback inhibition and herewith support segregation of figure from ground (e.g., Erb and Aertsen, 1993; Grossberg and Pessoa, 1998).

Object specific increases in population spike rates (multiple unit activity, MUA) were indeed recently reported for neurons representing object surfaces relative to those representing background in monkey V1 (Lamme, 1995; Zipser et al., 1996). Objects were defined by the contrast among textures consisting of the same elements inside and outside the object except for the value of one feature, as for example local line orientation. Object specific rate increases disappeared under anesthesia, which was taken as indication of top-down influence (Lamme et al., 1998). Analysis of the temporal development of spike rate modulations suggested that figure–ground segregation may be initiated from the detection of contours, followed by the filling-in of the surfaces between these contours to render them visible (Lamme et al., 1999).

Feature grouping based on synchronized gamma-oscillations. Alternatively, figure–ground segregation may be based on the grouping of surface elements that are similar and adjacent, so that contours are identified as the borders between differing elements. This coding possibility is supported by the observation of cortical oscillations in the gamma range (35–90 Hz) in the visual cortex of cats (Eckhorn et al., 1988; Gray et al., 1989) and monkeys (Kreiter and Singer, 1992; Eckhorn et al., 1993a) because they occur in several aspects according to Gestalt criteria of feature grouping (reviews in: Roelfsema et al., 1996; Eckhorn, 1999; Gray, 1999). In this concept and related models, neurons representing features of the same object, couple

their activities transiently by forming synchronized assemblies (Eckhorn et al., 1988; Gray et al., 1989; Eckhorn et al., 1990; Engel et al., 1991a; reviews in: Gray, 1999; Eckhorn, 1999). Segregation of figure from ground can in this concept either be coded by phase shifted (e.g., Horn et al., 1991; Stöcker et al., 1996) or decoupled γ-oscillations among object and background representations.

Contradictory findings of figure–ground segregation based on signal decoupling. In one type of experiment, investigating scene segmentation, neurons with overlapping receptive fields of different orientation preference were stimulated by either a single moving light bar or by a superposition of two light bars of different orientation and movement direction (cat V1: Engel et al., 1991a; monkey MT: Kreiter and Singer, 1996). While a single bar was reported to induce γ-oscillations that were coupled in many recordings, the two bar situation reduced coupling and often generated two separate groups of neurons with internally correlated oscillations.

In a second type of experiment, neurons with separate receptive fields were recorded and stimulated either by a single moving stimulus (light bar or grating) or by two stimuli with a gap among them moving either in the same or in antagonistic directions (cat V1: Gray et al., 1989; V2: Brosch et al., 1997). For the single stimulus condition these studies also reported strongly correlated γ-oscillations for many of the activated neurons, and in the two stimulus situations oscillations with reduced correlation. Hence, even though both experiments did not directly test the classical figure–ground situation, they came to related results suggesting that segregation is supported by decoupling the activities of an object against that of its background representation.

However, a recent investigation in monkey V1 seems to challenge the hypothesis of figure–ground segregation by decoupling (Lamme and Spekreijse, 1998). In this study, the relationship between the correlation of pair recordings and the perceptual organization of the scene was in many cases not consistent with the hypothesis of feature-linking-by-synchronization. In their experiments various scenes were presented, consisting of a figure on a background composed of similar texture elements. In particular, the strength of multiple unit (MUA) cross-correlations between pairs of recording sites often were not significantly different, when the sites represented elements of the same scene segment compared to when they represented figure and ground. Consequently, the authors concluded that synchrony in V1 does not reflect the binding of features.

Our investigation. In our present investigation we primarily want to clarify the seemingly contradictory results of coupling and decoupling among representations of two different segments of a scene, in particular the effect exerted by a contour defined by textures. Other than yet investigated coupling mechanisms may play a role in figure–ground coding. They should become visible in the modulation of other types of coupling measures or in different frequency ranges. In our investigation we therefore analysed several signal components and measures, including cross-correlation, coherence in the gamma (γ: 35–90 Hz) and low-frequency (1–20 Hz) ranges of multiple unit activity and local field potentials and time courses of spike rate modulations. Monkeys were shown a figure–ground stimulus consisting of object and background regions of as much identical features as possible. We chose a grating texture of equal contrast, spatial frequency, and orientation, in which the object was exclusively defined by an offset in the grating's spatial phase within a quadratic region. In contradiction to the conclusions of Lamme and Spekreijse (1998), our results from V1 suggest that figure–ground segregation is supported by desynchronization. We demonstrate this by a systematic strong decoupling across the object's contour yielded by a substantial reduction in coherence of γ-LFPs between inside and outside of an object's representation in two monkeys. Additionally, an object specific rate enhancement was present in one monkey, confirming previous observations partly (Lamme, 1995; Lamme et al., 1998, 1999).

2.2 Methods

2.2.1 Visual stimulation and behavioural task

Visual stimulation. Visual stimuli were applied using a 21-inch computer monitor with 98 Hz frame rate and 800×600 pixel resolution at a distance of 125 cm in front of the monkey, covering $9.1° \times 6.8°$ of the monkey's visual field. The stimulus consisted of a sinusoidal grating extending over the whole screen. Within this area a square object was defined by a $2.75° \times 2.75°$ part of the grating shifted in spatial phase (Fig. 2.1; grating luminance: 0.6–58.0 cd/m^2 ; Michelson contrast 98 %; luminance of the homogenous grey screen and the surrounding walls were adjusted to the average luminance of the grating of 6 cd/m^2, considering a logarithmic scale). The spatial frequency of the grating was kept constant within each recording session while it could vary between sessions. We used frequencies of 1–5 cycles/°,
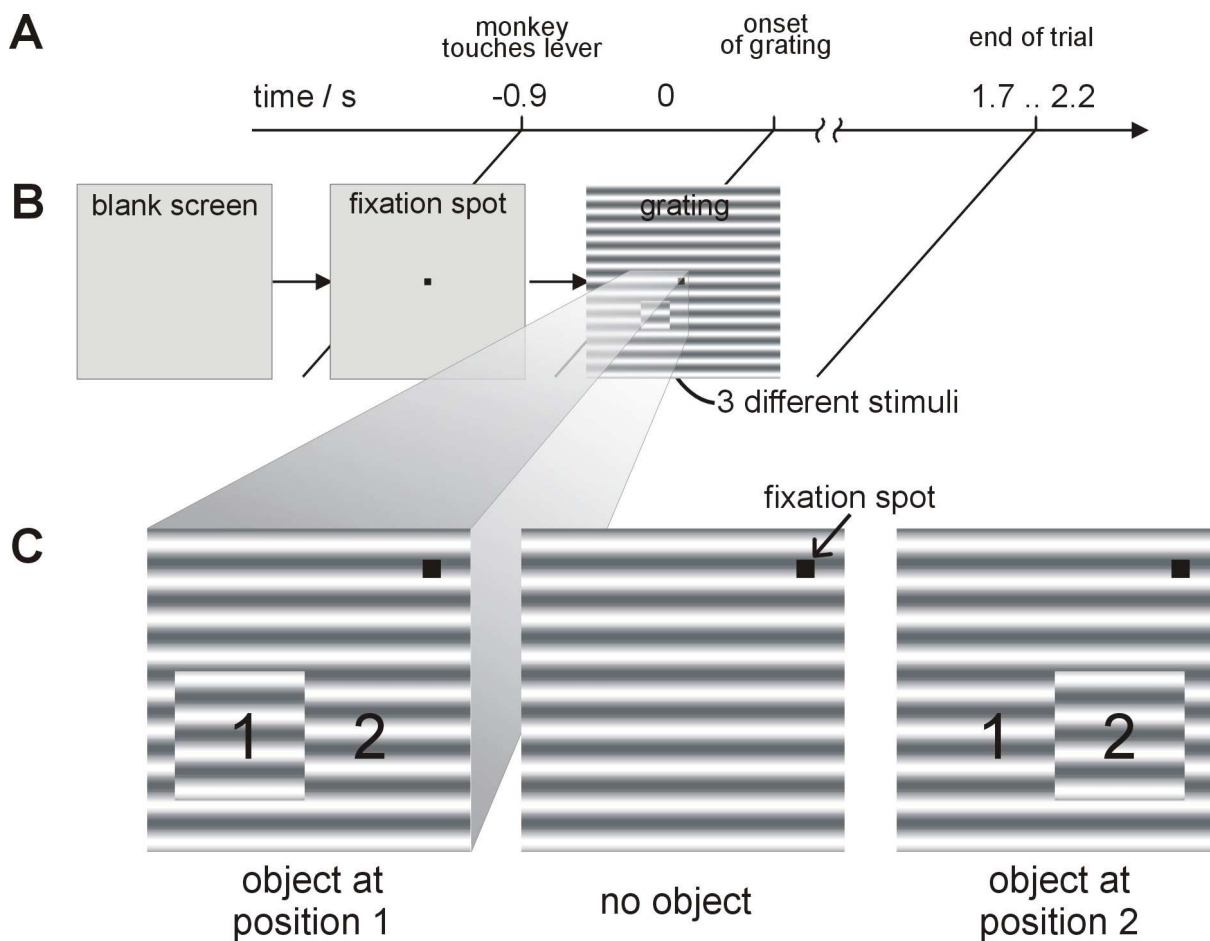
Figure 2.1 Time course of stimulus presentation and behavioral control of a single trial. **A:** Succession of events. **B:** Sequence of stimulation. **C:** Enlargement of the stimulation area containing the object and the fixation spot (black square). Note that the object in positions 1 and 2 is mirrored at one of its vertical edges.

which guaranteed adequate activation at the recording eccentricities chosen in V1. Within blocks of trials the orientation of the grating was kept constant, but it was varied between blocks to allow for different arrangements of the classical receptive fields relative to the object (Fig. 2.2, and see below).

Time course of stimulus presentation and behavioural control. The monkey fixated a green spot ($6.5' \times 6.5'$ visual angle; $39\,\mathrm{cd/m^2}$), which appeared at the center of the screen when he indicated his readiness for the next trial by touching a lever (Fig. 2.1A, B). This fixation spot remained visible during the trial. The trial was ended immediately when the monkey released the lever or failed to keep fixation within $0.8° \times 0.8°$ visual angle. Eye movements were controlled by an infra-red camera system with $225\,\mathrm{Hz}$ frame rate and a resolution of $0.1°$, developed in our group. Unless the trial was aborted early, the fixation spot dimmed to $23\,\mathrm{cd/m^2}$ at its end. Subsequently the monkey had maximally $0.65\,\mathrm{s}$ to react by releasing the lever. He only was rewarded with water or juice when he did not interrupt the trial early and managed to

react to the dimming in time. These trials were the correct ones that were evaluated. The regular end of a trial was reached after a total of 2.6–3.1 s. This interval contained a 0.9 s prestimulus interval with a blank screen followed by the grating stimulus interval of random duration of 1.7–2.2 s (Fig. 2.1 A, B). Trials were separated by a pause of approximately 2 s duration.

2.2.2 Data aquisition and analysis

Preparation and recording. Experiments were performed on 2 male rhesus monkeys (Macaca mulatta) aged 7 (monkey H) and 12 years (monkey S). Preparation and recording were carried out as reported elsewhere (Eckhorn et al., 1993a; Frien et al., 1994) in accordance with German laws of animal maintenance and experimentation and the guidelines published in the NIH *Guide for the Care and Use of Laboratory Animals* (NIH publication no. 86–23, revised 1987). In short, under deep barbiturate anesthesia three stainless steel bolts were affixed to the skull for painless head fixation during the recording sessions, in addition to a stainless steel chamber of 8 mm outer diameter giving access to primary visual cortex (V1) through the intact dura. After a few days of recovery recording sessions started. In each session 7 linearly arranged quartz-isolated platinum-tungsten fiber-microelectrodes with 500 or 750 μm inter-electrode spacing (Eckhorn and Thomas, 1993) were inserted individually into the upper layers 2 and 3 of V1 and neural activities were recorded extracellularly in parallel from all electrodes.

Data aquisition. Two types of signals were extracted from the raw broad-band signal (1 Hz to 10 kHz). Multiple unit activity (**MUA**) was extracted by band-passing (1–10 kHz; 18 db/oct), full-wave rectification and subsequent low-pass filtering (140 Hz; 18 db/oct), yielding an amplitude-weighted measure of population spike activity near the electrode tip without rejecting low amplitude spikes. The mean MUA level during prestimulus recording (blank grey screen with fixation spot) was subtracted from the following response epochs. Second, local field potentials (**LFP**) were obtained by band-passing from 1 Hz to 120 Hz. Both analog signals (MUA, LFP) were sampled at a rate of 500 Hz.

Mapping of classical receptive-fields. Position and extent of the classical receptive fields were measured with the RF-cinematogram method (Eckhorn et al., 1993b). For this, a white spot of Gaussian luminance distribution changed its location pseudo-randomly every 30.6 ms on a hexagonal grid within a rectangular frame containing all pre-estimated receptive field positions. Cross-correlation between stimulus position and responses yields a time-resolved pro-
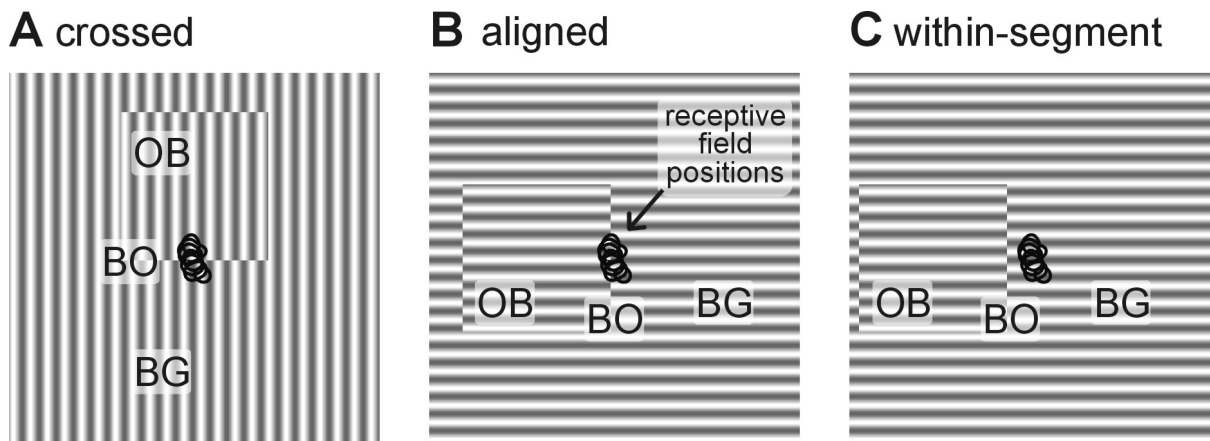
A crossed

B aligned

C within-segment

Figure 2.2 Three different arrangements of the classical receptive fields relative to figure and ground. Recordings were made by a linearly arranged array of 7 µ-electrodes. **A:** Receptive field array across the contour, **B:** along the contour, **C:** parallel to the contour, i.e., within a scene segment (average receptive field distance from the contour is the same as that of the outer receptive fields in A). Note that the object positions but not those of the receptive fields have been changed during the same recording session. (OB: object, BO: figure–ground border, BG: background).

jection of the neuron's spatial response profile in visual space, i.e., the temporal development of the spatial aspects of the receptive field structure. To extract the receptive field position and extent, the temporal dynamics were ignored. Receptive field measures were obtained by averaging over the epoch of the strongest response (50–120 ms post-stimulus). This raw receptive field profile was spatially interpolated and the contour of the 71 % level was considered the receptive field contour. Hence, its center of mass gives the receptive field center, the average diameter a measure for its size which is close to that obtained with the minimum response method (Barlow, 1972). In contrast to methods using a moving light bar as stimulus, this procedure avoids the uncertainty in reconstructing the receptive field position from the bar movement due to delay differences for stimulus movement directions with differences in response strength. Instead, a moving light bar of low velocity (1.5°/s) was used to obtain the orientation characteristics of the recording positions. Eccentricities of receptive field centers were parafoveally between 0.9° and 2.1° with extents varying from 0.2° to 0.6°. Corresponding LFP receptive fields were on average 30 % larger in diameter (the sizes for MUA and LFP were determined with the same criteria and from the same raw recordings).

Classification of receptive field positions. The visual stimulus consisted of a rectangular object defined by texture contrast. Simplifying the situation, we defined 3 classes of receptive field locations relative to the area of the object by grouping them with respect to the location of their receptive fields (Fig. 2.2). Each recording position belongs to either the object's surface (**OB**), the object's contour ("border": **BO**) or the background (**BG**). This classification

was made on the basis of MUA receptive fields obtained by the RF-cinematogram method. Only channels with MUA of sufficient signal to noise ratio are considered in our analyses. Whenever the MUA receptive field contour at a given recording position intersects the object's border, it belongs to the contour representation (BO). Otherwise it is part of the object's surface (OB) or the background (BG), depending on which side of the contour it lies.

PSTHs, spectral analysis, and coupling measures. Different measures were calculated in our analyses in the temporal and spectral domains: peristimulus time histograms (PSTHs), power spectra, cross-correlation, and coherence functions. To estimate the stimulus-locked components of the signals and their coupling strength a shift-predictor (Perkel et al., 1967) was calculated for power spectra and cross-correlation functions. Modulations of these measures due to different stimulus conditions were calculated for the stimulus specific portions of the responses, which means the pre-stimulus (baseline) level of each measure was subtracted before calculating relative changes. For comparability with previous work modulations of cross-correlation coefficients were additionally calculated with subtraction of the shift-predictor (without subtracting pre-stimulus correlation).

Except for PSTHs, a sliding window technique was used with an epoch length of 128 ms and a shift of 32 ms. After windowing the data by a boxcar function the mean value was subtracted for each epoch individually. For spectral measures epochs were then multiplied by a Hamming window before calculating a fast Fourier transform.

Coherences were calculated across the number of trials with absolutely identical stimulation (Bartlett-smoothing: Glaser and Ruchkin, 1976)[1] and sorted with respect to different stimulus parameters such as object position or spatial phase of the grating. Additional averaging, e.g., across different spatial phases of the grating or across different recording positions, was performed on these coherence values. As for all other normalized measures, this was done with Fisher-Z-transformed values.

The spectral resolution was restricted to 7.8 Hz due to the window length. Two frequency ranges were distinguished: the low-frequency range up to 20 Hz (bin 0: 0–4 Hz to bin 2: 12–20 Hz; note that for LFPs the lower band-pass limit is 1 Hz in our recordings) and the γ-range from 35 to 90 Hz (bin 5: 35–43 Hz to bin 11: 82–90 Hz). The latter range is determined by the typical stimulus specific increase of the average signal power within this range and is separated from the low-frequency range by a gap of low signal power around 30 Hz.

---

1   for a more formal description of the coherence measure see also Paragraph 4.2.2

Receptive field arrangements and object related modulations. To compensate for possible dependencies on contrast polarity of the neural activation at single recording positions, we used 3 different absolute spatial phases of the grating and averaged data across these conditions. The spatial phases of the object and background areas were varied independently. As a consequence, in one third of the trials the phase difference between object and background was zero and therefore no object but a continuous grating was visible (Fig. 2.1C center). With a phase difference and the object visible, it could appear at two positions mirrored at one edge (Fig. 2.1C). This ensured equal relative frequency for each recording position (not part of the contour representation) to be part of the object and the background representation while keeping the same distance to the contour. Variation of the spatial phases ensured on average locally identical stimulation in both conditions.

Comparison of the two conditions with different object positions allowed to extract object specific modulations like Lamme and colleagues did (e.g., Lamme, 1995). Comparison of the conditions with or without object allowed to extract modulations specific for texture segmentation.

To analyse data with respect to both kinds of modulation we made the relevant edge of the object, i.e., the one the object's position was mirrored at, intersect the linear array of receptive fields halfway. We call this condition *crossed* arrangement (Fig. 2.2A). Additionally, we used arrangements with all receptive fields lying at the object's contour (*aligned* arrangement; Fig. 2.2B), or parallel to it with the same offset as the outer receptive fields in the crossed arrangement had (*within-segment* arrangement; Fig. 2.2C).

Stimuli for different receptive field arrangements during the same recording session were presented blockwise, while variation of all other stimulus parameters was done pseudo-randomly within blocks of trials.

2.3 Results

2.3.1 Analysed data

In 19 recording sessions data sets from 133 recording positions were obtained. After visual inspection and rejection of faulty recording positions (due to broken electrode tips and other artefacts) complete recordings from 54 of 56 positions (monkey H) and 68 of 77 positions (monkey S) were used for further analysis.

Pairwise coupling of neural responses was characterized by the coherence in the low-frequency and γ-range and cross-correlation coefficients at zero time lag (baseline or shift-predictor subtracted). The four coupling measures were calculated for MUA and LFP for the first five post-stimulus epochs in both monkeys.

We primarily addressed the question of reduced coupling between object and background representations, and for this condition LFP coherence in the γ-range is the only one of the determined coupling measures that shows this reduction at high significance in consecutive epochs. We therefore mainly present results on the basis of LFP γ-coherence. Comparisons with the other coupling measures are made in Results and Discussion.

2.3.2 Stimulus-related modulation of coherence: within-segment condition

Within the representation of the same scene segment (object, background or whole field) stimulation induced particularly high values of LFP coherence in the γ-range (35–90 Hz). Figure 2.4 shows the average temporal development of coherence during the first 254 ms post-stimulus when both recording sites are part of the object's surface representation (solid curve). During the early post-stimulus epoch (0–126 ms) LFP coherence is elevated within a broad spectral range compared to the pre-stimulus epoch (dotted curve). This stimulus specific increase in LFP coherence is especially present above 30 Hz. With advancing time the range of elevation shrinks to a band of about 40 to 80 Hz. MUA shows on average very weak stimulus specific γ-coherence (Fig. 2.4), even though MUA is on average as vigorous in its early response and as sustained in its later response phases as LFP (Fig. 2.3). The pattern of coherence modulation in Figure 2.4 is identical when the same recording sites are part of the background representation and is almost the same with stimulation by the continuous grating (dashed curve).

In principle, the within-segment condition can be analysed in the crossed receptive field arrangement (Fig. 2.2A) as well as in the within-segment receptive field arrangement

Figure 2.3 Average peri-stimulus time histograms (PSTHs) in response to the continuous grating stimulus pooled from both monkeys. The vigorous transient early responses and the sustained continuing activation of multiple unit activity (MUA) and local field potentials (LFP) were separated from the same raw data. They indicate that both types of signals were, on average, equally well driven by the grating stimuli (switched on at time zero). Note that we calculated root-mean-square values for LFP in this depiction.

(Fig. 2.2C). In the latter condition electrode pair distances range from 1 to 6 inter-electrode spacings (of 500 μm or 750 μm each). In the crossed receptive field arrangement maximally 3 receptive fields were lying at one side of the contour. This means that only recordings from distances of 1 or 2 electrode spacings contribute to averages in this condition. Nevertheless, the average time courses of coherence look qualitatively the same in both arrangements.

2.3.3 Stimulus-related decoupling of γ-activity among object and background representations

γ-coherence. The coupling behaviour between neural populations of the object and background representations is different in the continuous compared to the figure–ground condition, which was investigated in the crossed receptive field arrangement (Fig. 2.2A). Figure 2.5A shows a clear decoupling effect indicated by a substantial decrease of LFP γ-coherence. Strongest relative reduction occurs in both monkeys nearly in the same post-stimulus epoch (around 180 ms). For monkey H, LFP γ-coherence across the object's border nearly drops to the prestimulus level in the interval 128 to 254 ms post-stimulus. This occurs simultaneously with the object specific spike rate enhancement at the object's surface representation in this monkey (not shown). In monkey S the dissociation of the coupling on the continuous grating versus the coupling across the contour reaches its maximum about 30 ms earlier than in monkey H (96–222 ms post-stimulus).

During the early post-stimulus epochs (0–126 ms) relative differences in γ-coherence between the continuous and the figure–ground condition are small (H: 15 %, S: 10 %) and become large (H: 80 %, S: 78 %) during the later response epochs (H: 128–154 ms, S: 96–222 ms) which indicates that the stimulus-locked components of the early responses do not show contour related coherence reduction (see Discussion).

As data of both monkeys shows qualitatively the same effects we merged it (Fig. 2.5B, lower panels). Differences between stimulus conditions were tested separately for each coupling measure and epoch with the Wilcoxon rank-sum test. To compensate for error accumulation due to multiple testing, the alpha criterion was conservatively corrected with a factor of 40 (4 different coupling measures × 5 post-stimulus epochs × 2 signal types: MUA, LFP) due to Bonferoni, that is, an effective α=0.01 was tested with α'=0.00025. The main results of these tests are summarized in Table 1. The reduction in stimulus specific LFP γ-coherence ("stimulus specific" refers to the fact that the estimate for non-specific coherence during the prestimulus interval has been subtracted, cf. Methods) is highly significant for the three later
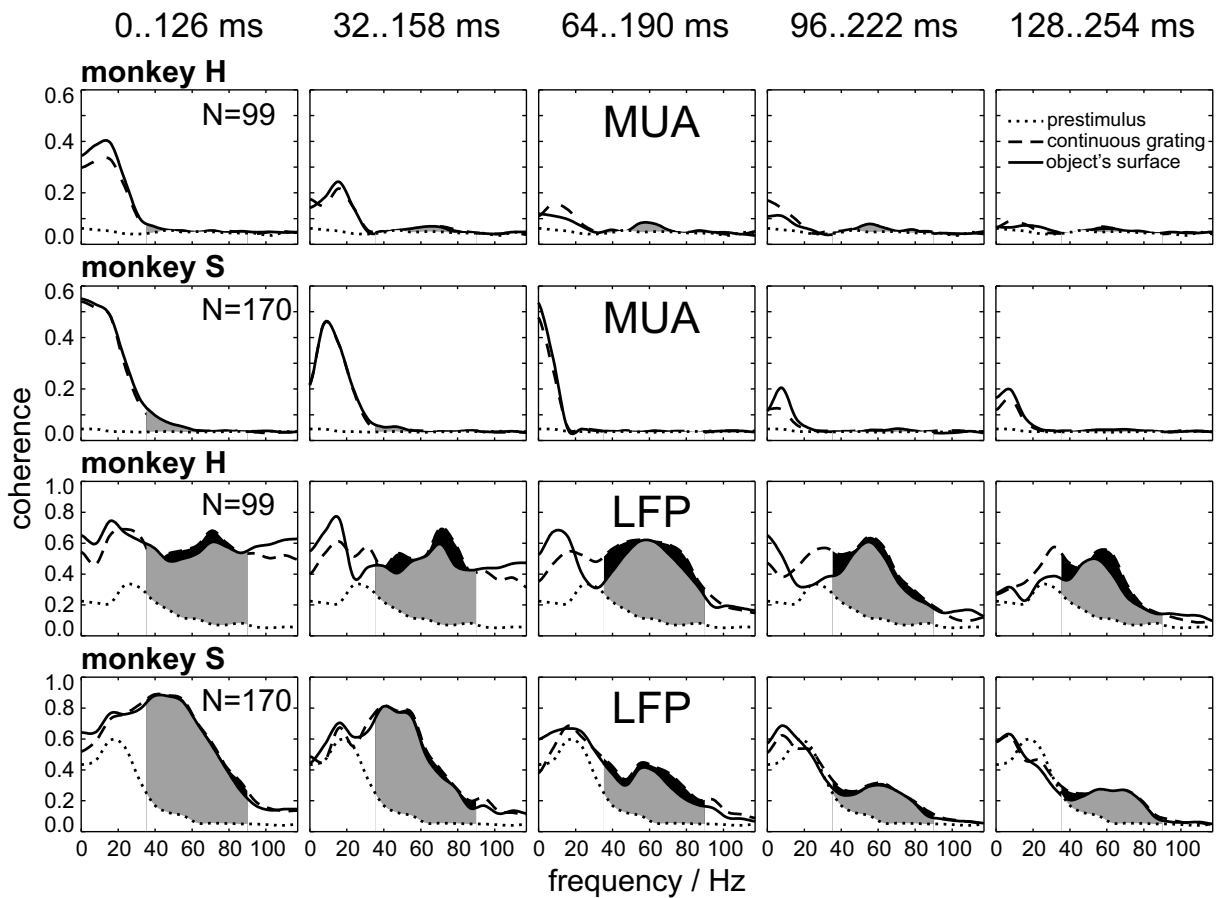
Figure 2.4 Post-stimulus temporal evolution of average MUA and LFP coherence in the within-segment condition (monkey H: N = 99, S: N = 170). Coherence in positions at background representation are nearly identical to those at object representations (continuous lines) and therefore not shown. Pre-stimulus values are included in all diagrams for comparison (dotted lines). Shaded areas indicate the γ-range (35–90 Hz). Gray shading: difference between pre-stimulus and within-segment coherence. Black shading: difference between continuous grating (dashed lines) and within-segment condition. The broad-band increase in LFP-coherence directly after stimulus-onset declines in the lapse of time particularly at low frequencies while LFP γ-coherence remains enhanced relative to prestimulus values. Note that the coherence for the condition of both receptive fields within the same segment compared to that of both at the continuous grating are very similar. Note also the different scales for MUA and LFP.

response epochs (64–254 ms post-stimulus) while this is not the case for the first two epochs (0–158 ms). There are no significant differences in coherence during any stimulus epoch for the low-frequency band, neither with LFP (Tab. 1 and Fig. 2.5B upper panels) nor with MUA, indicating that there are no consistent changes in coherence for low-frequency components among figure and background positions.

MUA cross-correlation. MUA cross-correlation analysis was calculated for comparison with previous work (Lamme and Spekreijse 1998). It reveales the same tendency for the reduction in coupling strength among figure and background positions as seen with γ-coherence of LFP, but less clearly (Tab. 1). The reduction in cross-correlation coefficient does not become highly

significant in any of the post-stimulus epochs when the pre-stimulus correlation is subtracted (while the shift-predictor is not subtracted), and it does become highly significant only during a single epoch (128–254 ms) when the shift-predictor is subtracted (while the pre-stimulus correlation is not subtracted).

2.3.4 Reduction in γ-coherence near the contour representation: control experiments

The decoupling across the object's contour would not be of greater interest if decoupling was induced by the contour equally well among within-segment and among across-contour positions. We therefore used the alternative within-segment receptive field arrangement by rotating the figure–ground stimulus while keeping the identical recording sites as control (Fig. 2.2 C). In this condition approximately the same distance of the contour to the receptive fields of both sites was realized as in the crossed receptive field arrangement. A major difference was the collocation of the receptive fields lying in the crossed-condition on either side of the contour and in the within-segment arrangement on the same side. Here, differing from Figure 2.4, data of the within-segment arrangement only includes those electrode pairs also contributing to data in the crossed arrangement (N = 34 for monkey S; N = 14 for monkey H), which are those positions in the crossed arrangement not lying on the object's contour.

In this control condition the reduction in γ-coherence is relatively low (<20 %) compared to the case when the receptive fields lie on different sides of the contour (80 %; see Fig. 2.6). This demonstrates that decoupling across the contour representation is due to the intersecting character of the contour and not due to its vicinity alone (see Discussion).

2.3.5 γ-power and -coherence along the contour representation

In search of potential mechanisms of decoupling across the contour, the coupling strength at the contour itself was analysed. This revealed lower values of LFP γ-coherence compared to the continuous grating and the within-segment condition during the post-stimulus epochs between 64 and 256 ms. Although the stimulus-related reduction is not as drastic as across the contour it is yet remarkable (about 40 %). However, stimulus specific LFP γ-amplitude density is also decreased at representations of the contour compared to that of the continuous grating: max. 35 % in monkey H and max. 40 % in monkey S (52–60 Hz; 64–190 ms).
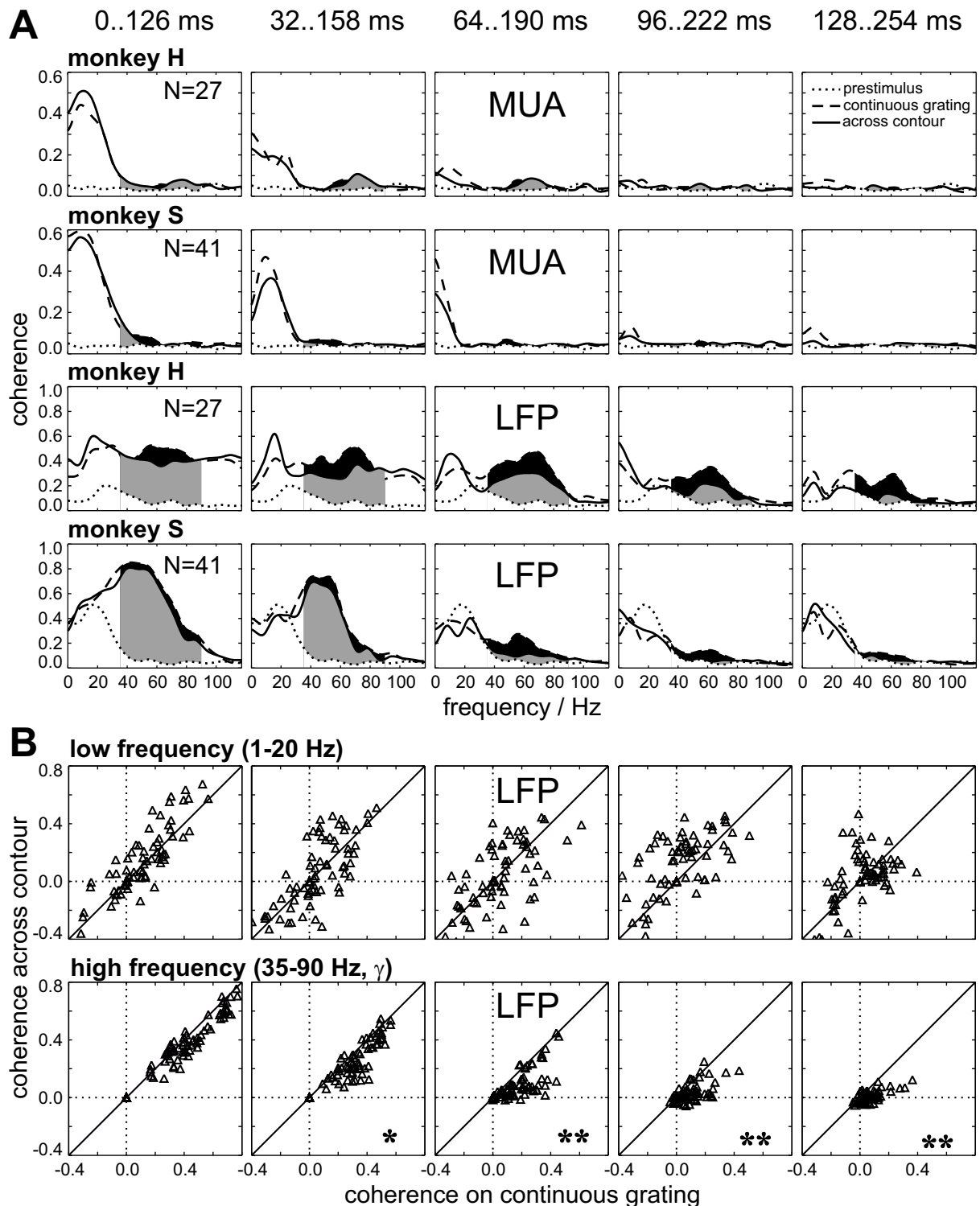
Figure 2.5 Post-stimulus temporal evolution of average MUA and LFP coherence in the across-contour condition (solid lines). **A:** Average coherence for monkeys H (N = 27) and S (N = 41). Pre-stimulus values are included in all diagrams for comparison (dotted lines). Shaded areas indicate the γ-ange (35–90 Hz). Gray shading: difference in coherence between pre-stimulus and across-contour conditions. Black shading indicates the difference between continuous grating (dashed lines) and across-contour conditions. Note the strong and broad-band increase in LFP coherence for both receptive field arrangements directly after stimulus-onset (particularly in monkey S). **B** Upper panels: Comparison of low-frequency LFP coherence (1–20 Hz) obtained with continuous grating (abscissa) and across-contour arrangement (ordinate); pooled data from both monkeys. Lower panels: Comparison of LFP γ-oherence with continuous and across-contour arrangement. Note that the gamma but not the low-frequency LFP coherence is highly significant reduced during the late three post-stimulus epochs in the crossed compared to the continuous condition.

2.3.6 Comparison of different coupling measures in all combinations of relative receptive field positions

To summarize the pattern of stimulus specific modulations in coupling Figure 2.6 compares relative values of different measures in all combinations of relative receptive field locations. Since pairs of recording sites with different average distances are pooled for the different groups of relative receptive field locations (OB*OB, OB*BO, etc.) we normalized the coupling values for each electrode pair to the corresponding value in the continuous grating condition before averaging. This enables direct comparison of stimulus-related modulations for the different situations Baseline or shift-predictor values were subtracted. For Figure 2.6 we only consider those recording pairs showing positive coupling values in the continuous grating condition after this subtraction, because we wanted to know the modulations which are specifically due to insertion of the object. In addition, results in this compilation (Fig. 2.6) are calculated in the optimal time interval for each monkey, i.e. 128 to 254 ms for monkey H and 96 to 222 ms for monkey S, respectively.

In Fig. 2.6A four levels of LFP γ-coherence can be distinguished. First, strongest coupling is present between groups of neurons representing the same scene segment when both

		coherence	
	epoch (ms)	*1-20 Hz*	*35-90 Hz*
MUA	*0 - 126*	0.392	0.458
	32 - 158	0.032	0.182
	64 - 190	0.014	0.275
	96 - 222	0.332	0.443
	128 - 256	0.0020	0.149
LFP	*0 - 126*	0.283	0.058
	32 - 158	0.441	0.0010 *
	64 - 190	0.474	0.0000002 **
	96 - 222	0.0053	0.0000007 **
	128 - 256	0.244	0.0000002 **

		cross-correlation	
	epoch (ms)	*minus baseline*	*minus shift-predictor*
MUA	*0 - 126*	0.318	0.149
	32 - 158	0.250	0.193
	64 - 190	0.027	0.017
	96 - 222	0.00029 *	0.0023
	128 - 256	0.00031 *	0.000011 **

Table 2.1 Test results (p-values) for the differences between the coupling strength (either coherence or cross-correlation) across the contour and the coupling strength in the situation with the continuous grating for identical receptive field positions (Wilcoxon rank-sum test, N = 68). Stars indicate levels of significance after Bonferoni correction (*: $p_{eff} < 0.05$, **: $p_{eff} < 0.01$; for details see text).
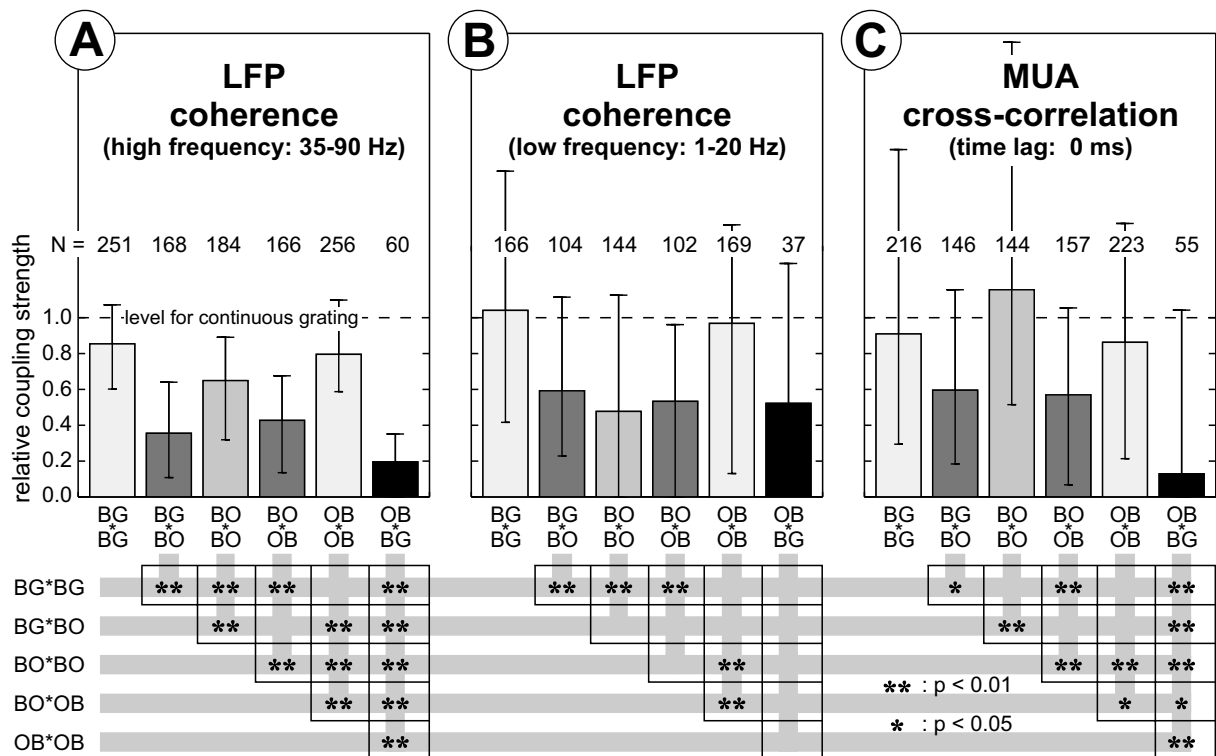
Figure 2.6 Coupling strength for different measures and different positions of the receptive fields relative to the figure–ground contour, normalized to the values obtained with the continuous grating. The bar heights represent the median ratio of coupling strength for the signal components in response to the figure–ground stimulus and the continuous grating (error bars indicate quartiles). Note that for coherence the baseline and for cross-correlations the shift-predictor was subtracted and negative values can occur. Therefore only those recording pairs were taken into account that showed positive coupling values in the continuous grating condition (their numbers N are plotted above each bar). The total numbers of recorded pairs for the different conditions are: 269 (BG*BG, OB*OB), 173 (BG*BO, BO*OB), 197 (BO*BO) and 68 (OB*BG). Data is merged from both monkeys and for each monkey we took the optimal post-stimulus epoch, 96–222 ms (S) and 128–254 ms (H), respectively. The matrices in the lower panel show corrected levels of significance for Wilcoxon rank-sum tests on differences between the coupling values for different receptive field locations. **A:** LFP γ-coherence. Due to the introduction of the figure–ground contour, within-segment situations (BG*BG and OB*OB) have slightly reduced LFP γ-coherence (about 20 %), while across-contour situations (BG*OB) show strong reductions in LFP γ-coherence (80 %). Intermediate reductions were obtained between border and one of the segments (BG*BO and BO*OB). **B:** LFP low-frequency coherence shows high variance (large error bars). Note, that the number of recording pairs with positive coupling values is much lower than for LFP γ-coherence. **C:** MUA cross-correlation results coarsly resemble those for LFP γ-coherences, but at much higher variance.

receptive fields lie on the object's surface (OB*OB) or on the background (BG*BG). Second, in the situation with both receptive fields on the object's border (BO*BO), coherence is lower compared to the within-segment condition (at about 65 % compared to continuous grating). Third, an intermediate coupling strength is observed for pair recordings where one receptive field belongs to the contour and the other to the object's surface (BO*OB) or to the background (BG*BO). Fourth, most important for the present investigation, weakest coupling (strongest decoupling) is present when one receptive field lies on the object's surface and the other on the background (OB*BG). Except the comparisons BG*BG versus OB*OB and

BG*BO versus BO*OB all differences in LFP γ-coherence are highly significant (Wilcoxon, p<0.01. See significance matrix in Fig. 2.6).

For low-frequency LFP many recording pairs do not show a stimulus-specific increase in coupling at all, as one can already see from Fig. 2.5B. There are significant differences in coupling modulations only between locations including the object's border versus those lying completely inside or outside the object representation. For figure–ground segregation the across contour condition (OB*BG) is relevant, but this shows no significant differences with respect to all other receptive field locations (right column of significance matrix in Fig. 2.6B). The modulations of MUA cross-correlations (Fig. 2.6C) coarsely resemble those of LFP γ-coherences, but at much higher variance (indicated by the large error bars) and therefore with significant results only in fewer combinations of recording positions. However, the reduction in coupling among figure and ground locations are highly significant compared to most of the other combinations.

2.3.7 Stimulus-related modulations in spike rate

Our figure–ground stimulus design has a lot in common with that of previous studies (Lamme, 1995; Zipser et al., 1996; Lamme et al., 1998) which demonstrated a late increase in spike population activity (MUA) within the representational area of an object's surface compared to that of the background in V1 of awake monkeys. Such object specific modulations in MUA are also clearly present in our monkey H between 150 and 260 ms post-stimulus. However, in monkey S the same analysis does not reveal any significant difference (data not shown).

2.4 Discussion

We demonstrate that the coherence of late LFP γ-components is strongly and significantly reduced across a contour's representation. This effect is also visible in MUA cross-correlations during late epochs but at much higher variance. However, γ-coupling is on average almost not influenced during the early response phase, when stimulus-locked components dominate, and this is also true for low-frequency coherences during the entire response epoch (LFP and MUA). Hence, our results suggest that particularly the decoupling of fast signal components is suitable for supporting figure–ground segregation.

2.4.1 Contradictory results from other work?

While our data reliably demonstrates substantial reduction of coupling across the representation of a texture contour compared to the continuous texture this was not the case in the work of Lamme and Spekreijse (1998) under similar stimulus conditions. In their experiment decoupling across contours is present only in a fraction of visual situations. This led them to the conclusion that scene segmentation is not supported by decoupling of signals at the contour representation. How can these differences be explained?

While we obtained our clearest results on the basis of LFPs, those of Lamme and Spekreijse are based on MUA. Evaluation of our calculations of MUA coherence showed no effect at all, while cross-correlations revealed highly significant modulations only during a short epoch if stimulus-locked components were subtracted (without subtraction of base line correlations. However, if base line correlations were subtracted (without subtraction of the shift-predictor) no significant differences in modulation were found. This means that MUA data is much less reliable in this task than LFP (Tab. 1), even though both types of signal components were extracted from the same raw data. In other words, the results by Lamme and Spekreijse are not really contradictory to ours, but differences are probably introduced by our use of LFPs, which seem to be more sensitive for effects of synchronization in visual cortex than MUA. However, our conclusion challenges that of Lamme and Spekreijse, because the reduction in LFP coherence in our results is highly significant and reliable and is taken therefore as an indication that decoupling of high frequency components seems suitable for coding scene segmentation.

2.4.2 LFP- versus MUA-recordings

LFP provides a conservative measure for the effect of coherence reduction in the present experimental situation, because LFP is volume conducted over larger areas than spike activities (half height decline of about $500\,\mu m$ and $50\,\mu m$, respectively). This is also visible in our receptive field measurements which have on average 30 % larger diameters with LFP compared to MUA (same raw data; same criterion). Volume conduction counteracts the effect of coherence reduction across the contour representation the more the closer the recording positions are to the contour, because decoupled activity from the other side of the contour representation is partly picked up.

Several properties of MUA are disadvantageous for the evaluation of the coupling phenomena analysed here. First, MUA contains strong broad-band stochastic power, including the

γ-range, that is not correlated with visual stimulation. Second, these broad-band components are only correlated among recording positions within a short distance range in V1 (König et al. 1995). Both properties result in MUA coupling values declining much faster to noise level with increasing cortical distance than values of LFP coupling. Hence, at many pair recording positions MUA cross-correlation and coherence are at noise level, which means they do not relevantly contribute to grand averages. This argumentation is supported by comparison of LFP and MUA results obtained from the same raw data in our present investigation. While LFP shows reliable coupling this is less reliable for MUA.

The higher sensitivity of LFPs for cortical coupling is probably inherent in their generation. They reflect the superimposed postsynaptic activities of thousands of neurons where the independent stochastic activities are canceled out while the coordinated fluctuations (the synchronized components) survive (Mitzdorf 1985). Thus, LFPs are well suited for analyses of interactions among spatially well separated populations of cooperating neurons ($> 500 \, \mu m$: half decline of volume conducted field). In summary, results similar to ours may have been obtained if Lamme and Spekreijse had used LFP for their analyses.

2.4.3 Reduction of LFP γ-coherence across the contour is not a result of a general reduction in γ-coherence

If the presence of a contour generally reduced coupling strength on either side of its representation (and not only across it), segregation by decoupling would not work. However, our data demonstrates that this is not the case. When the recording positions (which were kept constant over all stimulus situations; Fig. 2.2) were all at the object's or the background's representation in parallel to the contour, the appearance of the object and its contour induced only small reductions in γ-coherence of LFP (20 %) compared to the situation with one receptive field inside and the other outside the object (80 %). The small coherence reductions, due to the neighbourhood of the contour, parallels a decrease of γ-power in this situation, but our data gives no hint whether and how these effects are related. Anyhow, the strong reduction in γ-coherence across the contour cannot be explained by the small effects parallel to it because the signals at both sides of the contour's representation largely keep their within-segment coherence.

2.4.4 Not only γ-oscillations but also non-rhythmic γ-activity is decoupled across the contour representation

In our recordings the strong reductions in LFP coherence across the contour representation are present over the entire γ-range (35–90 Hz), independent of the nature of the signals being oscillatory or non-rhythmic. This indicates that the mechanisms of decoupling do not operate selectively on rhythmic signals. In our LFP data always a mixture of rhythmic and stochastic components is present (the former characterized by a spectral peak) with different relative contributions depending on the monkey and the post-stimulus epoch. However, these changing contributions of fast oscillations to γ-activity do not influence the strength of decoupling. This is important because many visual stimuli do not induce γ-oscillations but cause a shift in spectral content from dominating low frequencies before stimulation, resembling an alpha EEG, to broad-band γ-components (Eckhorn et al. 1993a; Frien et al. 1994; Jürgens et al. 1999), according to a desynchronized EEG (Vijn et al. 1991). As the stimulus-locked LFP components of the fast response at all frequencies and the later stimulus-induced components in the low-frequency range (1–20 Hz; not phase-locked to stimulus-onset) are not decoupled by the introduction of a contour, we assume it is the stimulus-induced γ-activity of the later response that is suitable for supporting figure–ground segregation. Probably the same neural mechanism of lateral signal decoupling is operative for rhythmic and non-rhythmic activity, and in addition, specifically influences the high but not the low-frequency components. However, for a generalization of our assumptions a broad variety of scenes has to be tested in this respect in future work.

2.4.5 Coherence reduction is probably due to blocking of lateral coupling connections and not to feedback from higher visual areas

Strong coherence reduction (80%) is present among receptive field positions directly adjacent to either sides of the contour but not for pair positions at the same distance to the contour on a single side. As the receptive fields of V1 neurons are much smaller than those of neurons projecting back from V2, and in particular from V4, TE and TEO (Rockland and Drash, 1996) one would not expect such precisely confined decoupling driven by back projections representing larger receptive fields. Hence, the spatial confinement of the step in coherence to the contour representation cannot be explained by decoupling actions from higher centers.

Instead, converging feed forward connections from V1 to these higher visual areas establish the receptive field properties of their target cells which may substantially be driven by

the coherent signal components of their inputs (e.g., synchronized γ-oscillations). Hence, the back-projection will carry just these coherent components, resulting in correlations of V1 activities with signals of neurons having overlapping receptive fields in the back-projecting visual areas (Eckhorn et al., 1990; Engel et al., 1991b; Frien et al. 1994). The back-projection is therefore more likely to enhance coherent components in V1 across the broad regions of the receptive fields of higher areas than exerting a spatially confined reduction at the contour. In summary, it is probable that the lateral connections within V1 become disabled at positions of contour representations. The mechanism of decoupling can, however, not be discovered from our present data.

2.4.6 No coherence reduction of stimulus-locked components at any frequency

We call *stimulus-locked* those signal components that are phase-locked to stimulus-onset. They are extracted from the raw signal by stimulus-triggered averaging (via shift-predictor) and have their prominent power in the short delay transient response. In the present investigation inspection of coherence of stimulus-locked components among inside and outside representations of the object revealed no significant difference after insertion of the contour, neither in the low-frequency nor in the γ-range (even though strong stimulus-locked γ-components are present; Fig. 2.5). This again indicates that coherence reduction in stimulus-induced γ-activity is probably due to the blockade of lateral V1 connections by contour representations. Segregation based on differences in stimulus-locked signals at figure and ground representations in V1 would be extremely fast. A recent paper on figure–ground effects with texture stimuli discusses this possibility (Lamme et al., 1999). It is shown that the fast stimulus-locked components evoked by the texture elements and texture-definded contours are not capable of supporting segregation between figure and ground, but a late increase in spike rate beginning 100 ms after stimulus-onset is. The occurence of this object specific component corresponds temporally to the epoch where we find significant reduction of γ-coherence across the figure–ground contour. As this late γ-activation is not phase-locked to stimulus-onset, it suggests in correspondence to the results of Lamme and Spekreijse, that contributions to figure–ground segregation may well be supported by signal components that are not strongly driven by the stimulus but are considerably influenced by intrinsic mechanisms: the object specific increase in the late spike rate is possibly influenced by top down effects (Lamme et al. 1998) and the decoupling of late γ-activity by the lateral cortical network.

2.4.7 Relevance of receptive field properties for figure–ground segregation by decoupling

In our analysis, we concentrate exclusively on the positions of the receptive fields and do not include other tuning properties like orientation preferences. This is done because the total number of recording positions (N = 122) is not sufficient for selecting enough cases with similar tuning at each side of the contour in order to make convincing statistical comparisons. However, we argue that perceived figure–ground contrast defined by textures is mainly based on the relative positions of local features at contours, and less, for example, by the local orientation of texture elements. Hence, if figure–ground segregation is based on γ-decoupling across contour representations, our results should not strongly depend on the local coding properties of the participating neural populations. This has not been analysed to date.

However, some indirect hints on how neurons with different orientation preference contribute to the effects analysed here are given by a recent work in which the same continuous grating has been used (Frien and Eckhorn, 2000). γ-, but not low-frequency MUA coherence in V1 of awake monkeys is shown to be the stronger the more similar the orientation preferences are, and the better the grating stimulates the neurons (which is the case when the grating's orientation matches that of the receptive fields). But coupling is reliably present even among recording pairs with perpendicular orientation preferences, if the orientation of the stimulus activates both populations simultaneously. This means, positions with different orientation preferences and suboptimal match to the stimulus orientation also participate reliably in γ-coupling. Therefore, in our LFP recordings the populations matching the grating's orientation are likely to have contributed the dominant components to the high γ-coherence during stimulation, while positions with suboptimal match contribute less. However, populations not participating in this task may contaminate the LFP by their unrelated components of maintained activity which reduce coherence nonspecifically. In conclusion, the LFP coherence reduction in our present investigation is a conservative measure, which means we would expect higher coherence, and correspondingly larger coherence reductions, if we selected pair recordings with orientation preferences matching that of the grating.

We want to note for the present work that the orientation of the grating and the positions of the receptive fields were generally not changed with changes from object to continuous background presentations which ensured that the same populations of neurons contribute to coherence with and without intersection by the contour.

2.4.8 Conclusions and perspectives

Our experiment allows comparison of several coding aspects of object representation in monkey V1 in parallel. Contour and surface specific modulations of spike rate and γ-coherence of LFP show reliable characteristics appropriate for supporting figure–ground segregation, and thus, object coding. They probably reflect complementary mechanisms rather than being redundant. While spike rate enhancement at an object's surface representation in V1 is meant to be driven by top-down influences (Lamme et al. 1998), we assume that the decoupling effects among representations of both sides of a contour shown by us is a property of the lateral network in V1. The potential contributions of V1 to object coding are surely not sufficient for unique scene segmentation and object specification, particularly for complex scenes containing multiple objects. Apparently, additional visual structures and possibly other mechanisms are involved. In addition, future research should test whether reduction of high frequency coherence is specific for a broad variety of visual situations. As figure–ground segregation may be supported by attentional mechanisms as early as in V1 (Roelfsema et al., 1998; Vidyasagar, 1998), it would be interesting to study the combined action of rate enhancement at the surface representation of an attended object, as reported in recent work, and the γ-decoupling across its contour representation, as observed in our investigation.

3 Dichoptical Receptive Field Cinematogram

Simultaneous mapping of multiple binocular and both monocular receptive fields under natural vergence in awake animals

To determine binocular as well as left and right eye classical receptive fields (CRF) simultaneously we extended the common reverse correlation technique for mapping spatio-temporal CRFs to binocular plus dichoptical stimulation with three independent stochastically jumping spot stimuli presented in a modified Wheatstone stereoscope (dichoptical receptive field cinematogram). To minimize vergence mis-alignment and vergence–accomodation mis-match, when using the stereoscope with awake monkeys, we calibrated the setup by evaluating offsets between monocular and binocular CRF positions at multiple recording sites (16 electrodes) in parafoveal V1. We were able to thereby reduce spatial incongruency between binocular and left/right eye stimulation to ~0.025° visual angle. CRFs can be determined with the same algorithm on the basis of single- or multiple-unit spike activity (SUA, MUA), or local field potenials (LFP). We demonstrate effectiveness of the method in terms of short recording times with measures of ocular dominance based on MUA and LFP recordings. Due to the simultaneity of monocular and binocular CRF determination and the precise positional calibration the method is well suited to investigate left and right eye interaction in forming binocular CRFs during normal binocular viewing in awake animals.

3.1 Introduction

Stereo vision is an essential feature of visual perception. The interaction of neuronal activations via the left and right eye in constituting cortical binocular receptive fields plays a key role in deciphering the neuronal basis of stereo vision. Characterization of the classical receptive fields (**CRF**) with respect to ocular dominance (Macy et al., 1982; LeVay and Voigt, 1988) and disparity selectivity (LeVay and Voigt, 1988; Poggio et al., 1988; Poggio, 1995; Cumming and Parker, 1997) and its functional cortical organization are important issues in this context. Beyond this, spatio-temporal interaction of left and right eye monocular CRFs in forming binocular CRFs gains important insight into the mechanisms of stereo vision (Freeman and Ohzawa, 1990; Anzai et al.,1999a-c; Livingstone and Tsao, 1999; Prince et al., 2002a,b). Especially for the task of analysing the interaction of monocular CRFs dichoptical stimulation with correct eye vergence alignment and vergence–accomodation match are desirable.

Dichoptical stimulation. Two different kinds of dichoptical stimulators are commonly used to realize controllable stereo vision in neurophysiological setups. *Shutter systems* use time-multiplexed visual stimulation via a single optical track with switching stimuli alternately viewed by each eye by occlusion of the other. Alternatively, *stereoscopes* stimulate the eyes simultaneously via separate optical tracks with two pictures independently viewed, for example, via two mirrors in front of each eye (Wheatstone). Shutter systems have the advantage to allow natural eye vergence with appropriate accomodation – at least in the zero parallax / disparity condition – without any special effort, since left and right eye stimuli share the same location. However, time-multiplexing of the stimuli makes real simultaneity impossible and perceptually relies on flicker fusion instead. Technically this reduces the maximal available frame rate to half its value. Besides, perceptual crosstalk between the switched images (*ghosting*) due to slow monitor luminance decay (*afterglow*) or incomplete shutter occlusion can not fully be avoided. The same ghosting problem holds for *anaglyphic stereo imaging*, where dichoptical stimulation is realized by eye-specific color-filtering of stimuli composed of two different colors. Wheatstone-like stereoscopes do not suffer from this restrictions but require precise adjustments of either the eyes or the mirrors / screens to allow congruent view of the distinct left / right eye stimuli. With anaesthetized animals eye vergence in principle can passively be adjusted by the experimenter with an ophthalmoscope and miss-accomodation can be cor-

rected by appropriate lenses. Practically though, the vergence alignment is hard to keep and the precision of this procedure is unkown. With healthy awake animals or human subjects stimulus congruence is actively forced by the visual system through adjustment of the eye vergence, given that the left/right eye stimuli are in principle fuseable and not too far misaligned. However, in a stereoscopic setup which is not adequately calibrated, eye vergence may not represent the same state as during normal binocular vision and may not fit to the current eye accomodation. This is especially critical since it is unclear in how far vergence and/or fixation disparity influences cortical disparity tuning (Trotter et al., 1996; Cumming and Parker, 1999). In addition, when using incongruent left and right stimuli, as is the case in binocular rivalry paradigms, eye vergence will not automatically adjust unless using strong additional vergence forcing stimuli to allow the visual system to solve the stereo correspondence problem. Therefore, an as good as possible calibration of optical pathways in stereoscopic setups is desirable to allow adequate vergence and to minimize vergence-accomodation mismatch. In principle this is a deterministic geometrical problem, given that the absolute eye positions relative to the stereoscopic setup are precisely known. This, however, is not always practicable, especially in experiments with awake animals.

Our setup. We therefore developed a stereoscopic stimulation setup combined with a CRF mapping technique suitable for two purposes. Firstly, calibrating the stereoscopic setup *online*, with the animal in the acute recording position, via binocular and left/right eye monocular CRF mapping to allow dichoptical stimulation with as good as possible stimulus congruence. The left and right stimulation screens are thereby aligned relative to a third, binocularly viewed screen (Fig. 3.1A). This guarantees appropriate vergence and accomodation. In this sense the setup combines the advantages of shutter/anaglyphic sytems with those of classical stereoscopes. Secondly, the mapping technique allows investigation of left/right monocular interaction in formation of binocular CRFs (anaesthetized cat: Anzai et al.,1999a,b; awake monkey: Livingstone and Tsao, 1999; Prince et al., 2002a-c). With the calibrated setup this can be done in awake animals under *natural* binocular viewing conditions, where both monocular CRFs and the corresponding binocular CRFs are characterized at the same time.

The dichoptical CRF mapping (**DCRF**) technique is based on a sparse noise reverse correlation method (RF-cinematogram: Eckhorn et al., 1993b). In the DCRF binocular and left/right eye receptive fields are mapped simultaneously by means of independent stimulus functions on each screen and successive correlation of each stimulus function with the neu-

ronal response (cf. Freeman and Ohzawa (1990) for a pure dichoptical version). Simultaneity reduces recording time for the mapping procedure and avoids problems with instationarity in the recorded data. In addition, left/right eye interaction can be investigated while both eyes are equivalently stimulated. The method can be applied to various types of neuronal signals: we applied it to single unit activity (SUA), multiple unit activity (MUA) and local field potentials (LFP; see also Lohmann et al., 1988). We will present CRF evaluation including ocular dominance measures of multi-channel recordings (16 electrodes) in supragranular layers of primary visual cortex V1 of awake monkey.

3.2 Methods

3.2.1 Animal preparation and recording

Two male rhesus monkeys (Macaca mulatta) aged 9 and 14 were used for recording experiments. All procedures were carried out in accordance with German laws of animal maintenance and experimentation and the guidelines published in the NIH *Guide for the Care and Use of Laboratory Animals* (NIH publication no. 86–23, revised 1987). Shortly before the experimental sessions a plastic chamber of 10 mm o.d. was implanted under deep barbiturate anaesthesia to give access to primary visual cortex (V1) through the intact dura. Stainless steel head-posts for painless head fixation during recording sessions had been implanted in the same way years before.

In each recording session up to 16 quartz-isolated, platinum-tungsten fiber-microelectrodes were individually advanced into the cortex under acoustical and optical control of the recorded signals (Eckhorn and Thomas, 1993). The electrodes were arranged in a regular 4×4 array with 750 µm pitch. Two types of signals were retained from the raw broad-band signal (1 Hz to 10 kHz). Multiple unit activity (**MUA**) was extracted by band-passing (1–10 kHz; 18 db/oct), full-wave rectification and subsequent low-pass filtering (140 Hz; 18 db/oct). The mean MUA level during prestimulus recording (blank grey screen with fixation spot) was subtracted from the following response epochs. Second, local field potentials (**LFP**) were obtained by band-passing (1–120 Hz; 18 db/oct). Both analog signals (MUA, LFP) were sampled at a rate of 500 Hz.

3.2.2 Stereoscopic setup

We used a modified Wheatstone stereoscope for dichoptical and simultaneous binocular stimulation suitable for human and non-human subjects (Fig. 3.1A). In addition to the independent left and right eye stimulation in the classical setup, a third stimulation screen is used for binocular stimulation. Visibility of this screen is realized by replacing the reflecting mirrors of the classical setup with beam splitters (semi-transparent mirrors). Each eye's visual stimulus then consists of a superposition of the respective lateral and the central screen images. Since we invented this setup for a different purpose, the beam splitters in our setup do not match stimulus luminance along the central and lateral optical path. This means, for each eye the maximal luminance of a monocularly viewed lateral stimulus ($54\,cd/m^2$) is twice as high as of the binocularly viewed central stimulus. Apart from this, brightness and contrast settings of each screen were adjusted to yield an optimal match in luminance profiles for the
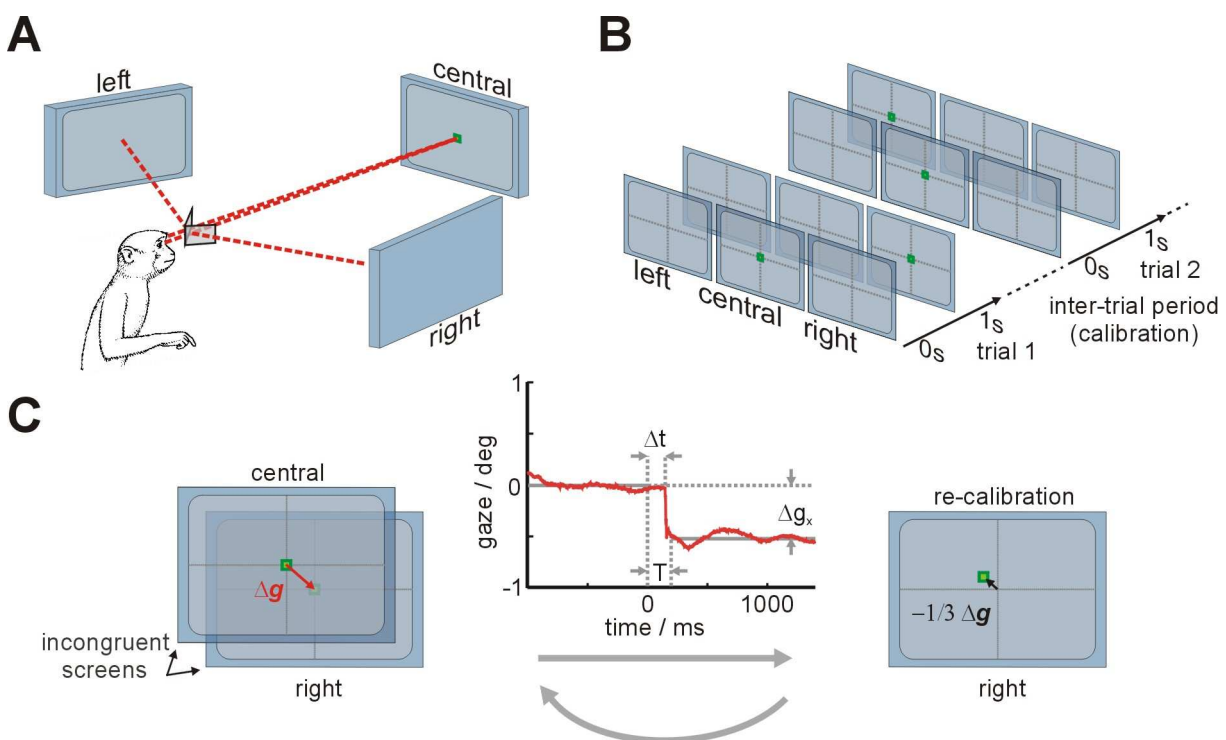
Figure 3.1 Dichoptical setup and pre-calibration. **A:** Modified Wheatstone-stereoscope for simultaneous binocular (central screen) and dichoptical (left / right) visual stimulation via semi-transparent mirrors (beam splitters). **B, C:** Pre-calibration procedure. In each trial the fixation spot is presented on the central screen for the 1st second, then switches to either of the lateral screens for about another second before dimming after a random interval (B). The monkeys' gaze direction is monitored with an infra-red eye-tracking system (225 Hz) calibrated on central screen coordinates. Changes in gaze direction Δg due to incongruent screen alignment (C left) are calculated from the eye position signal with 200 ms latency (T) after switching of the fixation spot (C middle; g_x shown only). The screen coordinates of the lateral screen currently on duty are re-calibrated trial-by-trial with a control-loop amplification of $-1/3$ (C right).

greyscale in use. The screens covered $18.2° \times 13.6°$ (resolution: 800×600 pixel, distance: 125 cm) of the subjects' visual field.

Each screen is driven by a separate PC but with identical graphics boards synchronized at 98 Hz frame rate. All 3 graphic boards share the same quartz clock signal of the central stimulation PC, i.e., the original quartz of the lateral graphic boards is removed. To synchronize video refresh the vertical synchronization trigger (VSYNC) of the central graphics board is awaited by the lateral graphics boards once at the beginning of the stimulation protocol. Well-functioning of this procedure was supervised by comparing the output currents of photodiodes located at corresponding positions of all three screens. Stimulation protocols of the three PCs are co-ordinated in real time via user-defined parallel port communication. When using greyscale stimuli exclusively, the technical effort for synchronizing video refresh on different screens could be reduced by using a colour-splitted RGB signal of a single graphics board (Prince et al., 2002b).

3.2.3 Positional calibration

Aim of the calibration procedure was to align each lateral stimulation screen to be congruent with the central screen. This eliminates vertical disparity between left and right and adjusts horizontal disparity to fit ocular accomodation for the distance of the central screen, where the behaviourally relevant fixation spot is presented. Positional calibration is performed in two steps. The first step is a pre-calibration based on a 3-screen fixation task with gaze control. The second step is based on the alignment of independently measured binocular and monocular left/right eye classical receptive fields (DCRF mapping). The first step is performed to minimize the to-be-scanned area in the second step, thereby optimising spatial resolution and hence precision of the positional calibration based on the CRF mapping. Alternatively, the DCRF-calibration could be performed repeatedly with successively adapted parameters.

In the pre-calibration task the monkeys had to fixate a small spot (0.1°) and indicate its dimming by lever release. During the trial the fixation spot switches from the central screen to the corresponding screen coordinate of one of the lateral screens (Fig. 3.1B). When the screens are mis-aligned this results in a change of gaze direction, which is monitored with a 225 Hz infra-red CCD camera system (Thomas Recording, Germany). Average gaze offsets Δg between stimulation via central and lateral screen are evaluated at the end of each trial, and the screen coordinates of the lateral screen that has currently been on duty is electronically corrected by a factor of $-0.3\Delta g$ (Fig. 3.1C). The control loop feedback amplification of -0.3 is

empirically choosen to optimize convergence of the procedure with the given inaccuracy of the monkey's gaze direction (cf. Results) and its determination (0.05°; Niessing, 1999), respectively. The process is auto-terminated after the absolute radial gaze offset under-went a limit value of ± 0.25 deg 10 times in consecution for each side.

For the CRF-based calibration we calculated the remaining screen mis-alignments from the average offset between CRF centres when determined binocularly versus monocularly for each side (for determining and parameterizing CRFs with the DCRF-method see below). This is an advanced version of the *reference-cell method* (Hubel and Wiesel, 1970; Ferster, 1981; Anzai et al., 1999a). It would be deficient to rely on an individual single cell CRF as reference for calibration since disparity tuning of individual neurons is partly formed by left and right eye CRFs with positional offset, and partly with phase-shifted CRF profiles (for most recent works on this issue: Anzai et al., 1999a; Prince et al., 2002a). Given the disparity tuning of the neuron in consideration is based on monocular CRFs shifted in position to one another, we would erroneously introduce the disparity preference of this particular neuron in the whole setup. Even MUA of a single channel would fail to overcome this problem, if neurons with similar disparity preference are locally clustered and the disparity mechanism is based on a positional shift. However this clustering seems to be only weakly the case in monkey V1 (LeVay and Voigt, 1988; Prince et al., 2002b). To obtain higher security against bias away from zero disparity calibration, we used the average offset of all available MUA CRF centres from simultaneous recordings of up to 16 electrodes (pitch 0.75 mm). Both, the left and the right screen are calibrated relatively to the central screen, not explicitly with respect to each other.

The CRFs have to satisfy certain criteria to be taken into consideration for the calibration procedure, including a sufficient signal-to-noise ratio (**SNR**; cf. below).

3.2.4 Receptive field mapping

The procedure and theoretical background of the standard backward averaging procedure for CRFs with single spot stimuli is portrayed in more detail in Eckhorn et al. (1993). We used a simplified and fully computerized version of this method (Fig. 3.2; formal description see Appendix). Stimuli are generated PC-based on a CRT and consisted of small bright spots with Gaussian luminance distribution. Stimuli are randomized only with respect to their successive positions in space within a 16×16 hexagonal grid with equal probability of occurrence at each grid node. Otherwise the spots are adapted in size to the expected CRF size (~1/2 diameter of

Figure 3.2 RF-cinematogram. **A:** In the computerized version of the RF-cinematogram (Eckhorn et al., 1993) a bright spot is successively presented at random positions of a 2-dimensional virtual stimulation grid for the dwell time of 3 video frames (30.6 ms). For the data presented here a 1-frame (10.2 ms) delay was inserted between successive spot presentations (not shown). The monkey has to perform a fixation task on a small green spot. **B:** At each grid node (n, m) the time course of stimulus onsets $T_{n,m}(t)$ represents the time of appearance of the spot at this node. (Note that the frequency of appearance of the spot at a single node in real data is more seldom than in the given schematic data). The physical stimulus function $S_{n,m}(t)$ at this location is given by the convolution of this time course with the temporal luminance profile of a single spot (cf. Appendix). The neuronal response function $R(t)$ is correlated with the time course of stimulus onsets at each grid node to give a time-dependent RF raw map $RF_{nm}(\tau)$ (B lower). Appearance of the spot within the RF area (n, m) yields high positive cross-correlation values (B left), while stimulation at outside RF positions (n', m') results in no stimulus-response correlation (B right). **C:** Keeping the latency information of the cross-correlation functions yields spatio-temporal RF properties (RF-cinematogram; C right). For the purpose of this study RF maps were averaged over the peak epoch of the cross-correlation function and spatially interpolated to extract RF position (C left; for details see Methods). All data shown is based on multi-unit activity (MUA) of a single channel. The procedure is equivalently repeated for all other simultaneously recorded channels by exchanging $R(t)$.

visible patch size for analysis of CRF position, not sub-structures) and in dwell time (3 video frames = 30.6 ms) to minimize total recording time against the background of a dwell time versus stimulus repetition trade-off (Schröder, 1996). To allow analysis of time- and amplitude-continuous signals like LFP and MUA (in the form of a time-continuous spike densitiy estimation) we applied *forward* stimulus–response correlation (Lohmann et al., 1988) instead of reverse correlation (spike-triggered averaging).

3.2.5 Dichoptical receptive field cinematogram

The dichoptical extension of the RF-cinematogram technique works with three stimulus sequences running in parallel but independently at congruent scanning areas on the three stimulation screens. This could be interpreted as expanding the stimulus dimensionality by the factor *stimulation pathway* consisting of three values: binocular, left or right monocular. But to save recording time, the pathway parameter is not randomized like the spot position. Instead, all three pathways are stimulated simultaneously in each stimulation time step, only that the sequences of spot positions among the pathways are independent from each other (Fig. 3.3A). This means, at every point in time there are three spots physically present, one viewed binocularly, the second through the left, and the third through the right eye only. Each spot occupies a random position within the scanned area. The stimulus functions at congruent positions but different pathways are in the same way statistically independent as stimulus functions at different positions within each single pathway. Evaluation of the DRFC therefore practically means calculating stimulus–response correlations for three times more stimulus positions, one third of belonging to each of the three pathways (Fig. 3.3B). For each pathway a spatio-tempo-

A

B

C
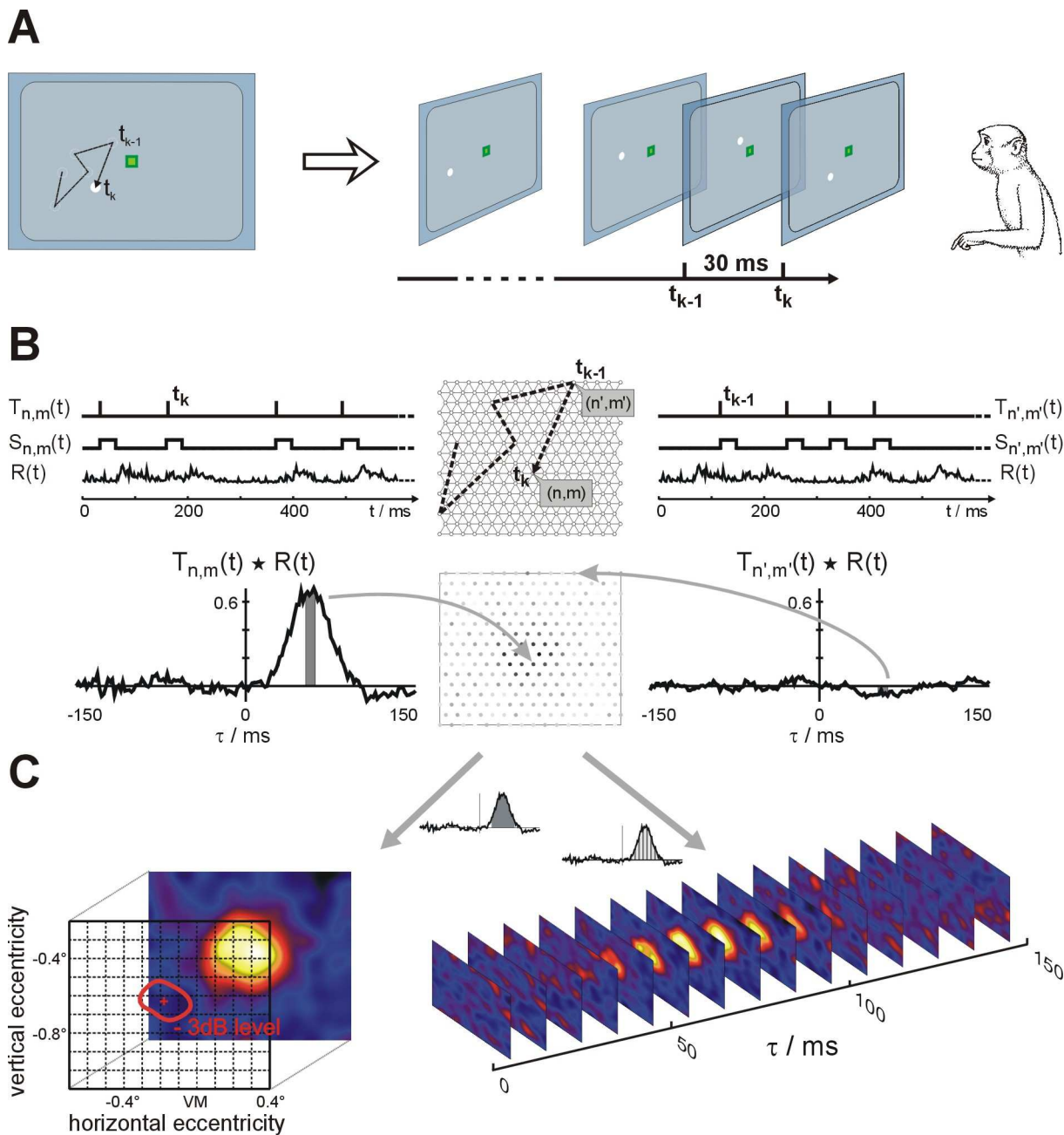
ral RF structure results, all representing the same visual field area but viewed differently: the binocular RF and the left and right eye monocular RFs (Fig. 3.3C).

## 3.2.6 RF parameterization

For the purpose of setup calibration the main target is to determine CRF positions. The dynamics of RF structures will play a secondary role here. The time-continuous neuronal signals are originally sampled at $500\,\text{Hz}$. To accelerate computation, data are compressed by a factor of 5 (to approximately the temporal resolution of the video refresh) by means of boxcar aver-

aging. Spike-density in single unit data is estimated by simply counting events falling into each bin, thereby bringing it into the same data format as the originally analogue data. Local field potentials are evaluated in a negativity-up manner according to EEG conventions. From there on, all types of signals are subjected to the same procedure. The mean signal amplitude across the entire recording time is subtracted. Correlation of the resulting neuronal response functions $R^e(t)$ with the stimulus-onset time courses (cf. equation A3.8 in Appendix) yields RF raw movies with the spatial resolution of the stimulus position grid and the temporal resolution of the resampled neuronal data (100 Hz; Fig. 3.2C right). Pseudo-shuffled reference data is constructed to estimate SNRs for defining data-based threshold conditions. This is done by time-reversing the neuronal response of each electrode over the term of the whole recording. This procedure sustains temporal characteristics of the data while sufficiently destroying stimulus–response correlations with negliable computational effort.

CRF positions are determined in a two-step procedure. In a first step those movie frames showing a positive stimulus–response correlation exceeding a pre-set threshold (time of interest: **TOI**) are extracted and averaged to yield a single RF raw map. In a second step the center of mass of a threshold-defined region of interest (**ROI**) is calculated on a spatially interpolated RF map. First step: To extract the time course of the neuronal response the CRF position within the spatial range of the raw frames has to be judged. With adequate parameter settings for visual stimulation more than one grid node will fall onto the area of the CRF. To make use of the resulting correlation in the response time courses associated with neighbouring grid nodes the raw maps are smoothed by means of next-neighbour averaging. This next-neighbour smoothing is only applied for the purpose of a preliminary CRF position judgement in the raw maps, not for further evaluation. The spatial coordinates of the maximum value across space and time in the smoothed frames is taken for the preliminary CRF position judge-

Figure 3.3 Dichoptical receptive field cinematogram. **A:** The RFC method is applied simultaneously via all 3 stimulation pathways (left / right monocular = left / right screen, binocular = central screen). The respective sequences of stimulus positions are statistically independent. The resulting RFs (cf. C) are underlaid for demonstrating the different times (t_k, t_{k+4}, t_{k+5}) when the distinct stimulus sequences hit the RF area. **B:** Neuronal response function $R(t)$ and stimulus functions $S^{L/B/R}_{nm}(t)$ for each pathway exemplary at a grid position (n, m) within the RF area. While left eye stimulation at the RF center results in a weak neuronal response (t_k), right eye and binocular stimulation elicits a stronger response (t_{k+4}, t_{k+5}). **C:** All stimulus onset time courses $T^{L/B/R}_{nm}(t)$ for the 3 optical pathways are independently correlated with $R(t)$ as is the case for a single pathway in the standard RFC method. For each recording site this results in 3 different RFs (left / right monocular and binocular) extracted from the same neuronal data $R(t)$. Maps in C are normalized to the maximum correlation value within the binocular map. This reveals right eye ocular dominance of the recording site by means of stronger stimulus-response correlation for right than for left eye stimulation. Note that a quantitative comparison of monocular and binocular response strength is difficult here since the physical contrast along the binocular pathway was weaker than along the monocular pathways. **D:** Respective monocular and binocular RFs of all 16 recording sites mapped simultaneously in this session. Electrodes were arranged in a cartesian 4×4 array with 750 μm nearest neighbour distance (see inset). Total recording time for this MUA example was about 120 s (30 trials á ~4 s).

ment. The time course of response at this maximum position is used to determine the TOI, which is the epoch of the first temporal response peak exceeding the predefined threshold by absolute value. The threshold is set to four times the spatio-temporal variance of the equally
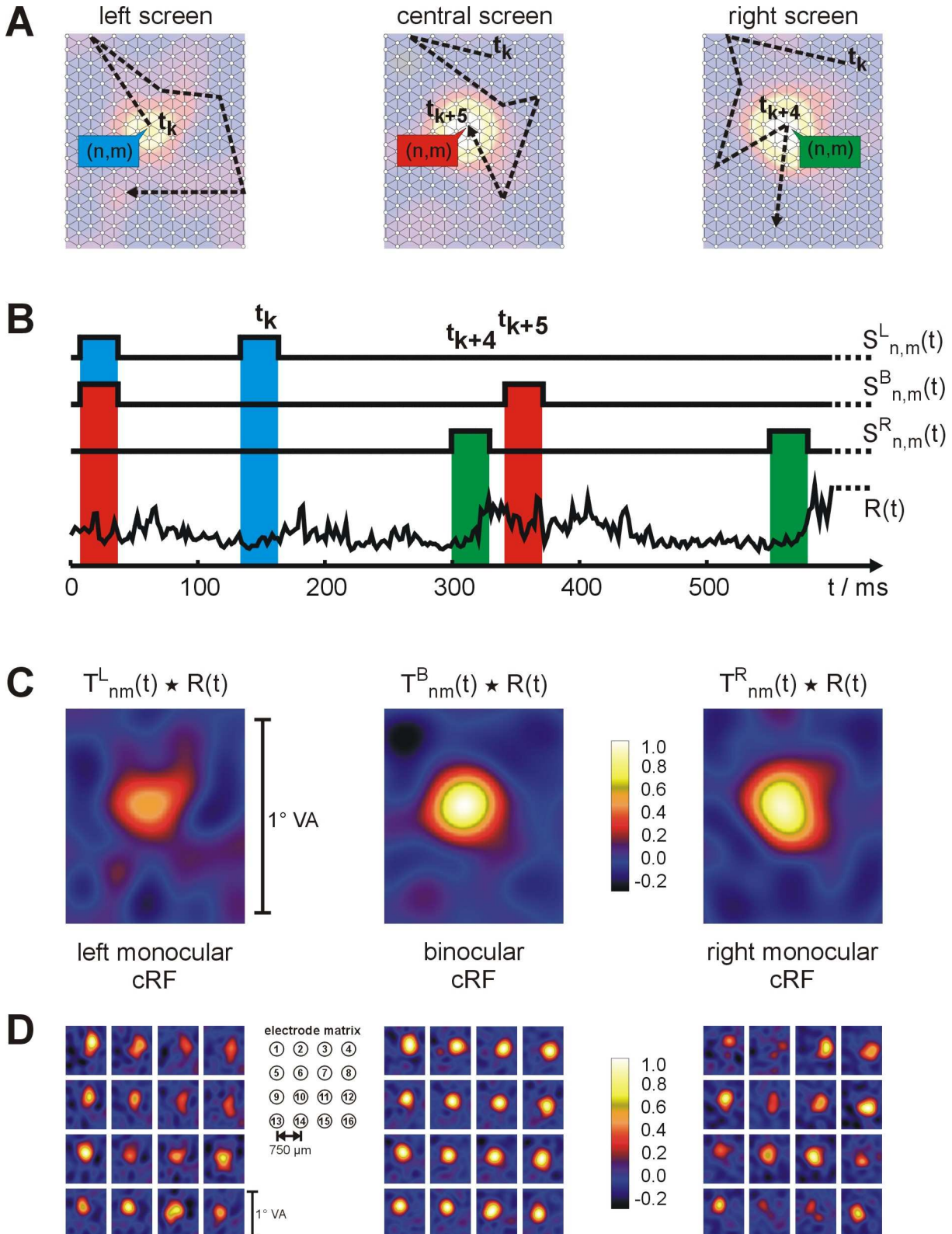
obtained reference maps (from the time-reversed data, cf. above). If no peak meets this *temporal SNR criterion*, a default TOI of 40–90 ms time lag is used instead. The TOI is determined on the basis of response to the binocular stimulation, and the identical TOI is used for evaluating the left and right monocular RFs. Spatial RF-parameterizations are solely based on RF maps averaged across the TOI (Fig. 3.2C left). Second step: RF maps are spatially interpolated to obtain RF maps with original pixel resolution of the scanned screen area. This resampling implies a transformation from the hexagonal sampling grid to the new cartesian grid, which can be done by upsampling and low-pass filtering via discrete fourier transformation, considering the shape of periodicity of the reciprocal hexagonal grid (Mersereau and Speake, 1983). For the data presented here, the system-theoretical ideal low-pass is replaced by a $\cos^2$-soft-edge cylindrical low-pass with cut-off frequency of half the value suggested by the Nyquist-theorem for the x-direction and one octave flank width (cf. definition of sampling grid in Appendix). We thereby reject spatially high-frequency noise, which eases peak extraction for determining the CRF position. A −3 dB amplitude threshold (71 % of maximum height within map) is used to define the ROI (Fig. 3.2C left). This ROI has to be a contiguous area to be accepted as a CRF, which practically means that no other peak within the map may reach the 71 % level (*spatial SNR* or *singularity criterion*). The position of the CRF is defined as the center of mass of the CRF peak, i.e., the amplitude weighted position-vector average within the ROI:

$$\vec{c} \;=\; 1/N_{ROI} \sum_{i=1}^{N_{ROI}} RF(\vec{r}_i)\vec{r}_i \qquad \vec{r}_i \in ROI \,. \tag{3.1}$$

This procedure would fail to consider CRF substructures appropriately, since it evaluates only peaks of one sign and ignores potential adjacent inhibitory regions. But such substructures can anyway not be seen in our MUA data from supragranular V1 used here for calibration.

Recording sites were characterized with respect to ocular dominance by means of an index (ODI) based on the receptive field mapping,

$$ODI = \frac{a_{ipsi}}{a_{ipsi} + a_{contra}} \;, \tag{3.2}$$

where $a_{ipsi/contra}$ denotes the responsiveness to the ipsi-/contralateral eye stimulation (Macy et al., 1982), here expressed as the amplitude of the respective CRFs. With strongly monocular CRFs, detection of the ROI within the weak eye's CRF map may fail. Therefore, the left/right CRF amplitude is calculated as average over the ROI of the binocular CRF (Fig. 3.8A):

$$a_{LIR} = \sum_{i=1}^{N_{ROI}} RF_{LIR}(\vec{r}_i) \qquad \vec{r}_i \in ROI_B. \tag{3.3}$$

When the binocular CRF is not available it is constructed as sum the of the two monocular maps for defining the common ROI.

3.3 Results

3.3.1 Positional pre-calibration

The typical convergence behaviour of the recursive pre-calibration procedure can be assessed from the example data given in Fig. 3.4. Within about 10 trials per side the procedure converges, when the initial screen incongruence amounts to about $2.5°$ visual angle. This initial offset is larger than the typical day-by-day inaccurancy of the setup, but used here for demonstration purposes. The precision based on the trial-by-trial feedback correction algorithm is not improved by more trials, as the variance (SD $= 0.10 - 0.14°$) of the eye position offsets for trials 15 to 30 demonstrate. This is due to the limited precision of the monkeys' fixation capabilities in this type of experiment: The residual eye movements (SD of gaze position) in the initial phase of the trials with binocular fixation spot on the central screen was smaller (average SD $= 0.063 \pm 0.003°$; N $=43$) than after the switch of the spot to the lateral screens with monocular viewing ($0.15 \pm 0.015°$). Original eye traces as recorded with our infra-red eye tracker during an adaptation performance of gaze direction are shown in the inlet of Fig 1C. Gaze directional response to the switching of the fixation spot normally occurs within less than 200 ms latency (174 ± 2.4 ms, N $=43$). The new eye positions were calculated as average from an offset time of 200 ms on, with the remaining data spanning about 800–1200 ms, depending on the exact trial length.

3.3.2 Reliability of single CRF localization

The major constraints for the precision of the DCRF calibration procedure are given by the residual uncertainty of the CRF localization. CRF parameters of the same recording channel determined in two successive blocks of trials within the same session match only with limited precision, as can be seen from the example in Fig. 3.5 (central column). We used the deviation of binocular CRF positions in two successive recording blocks to quantify reliability of CRF
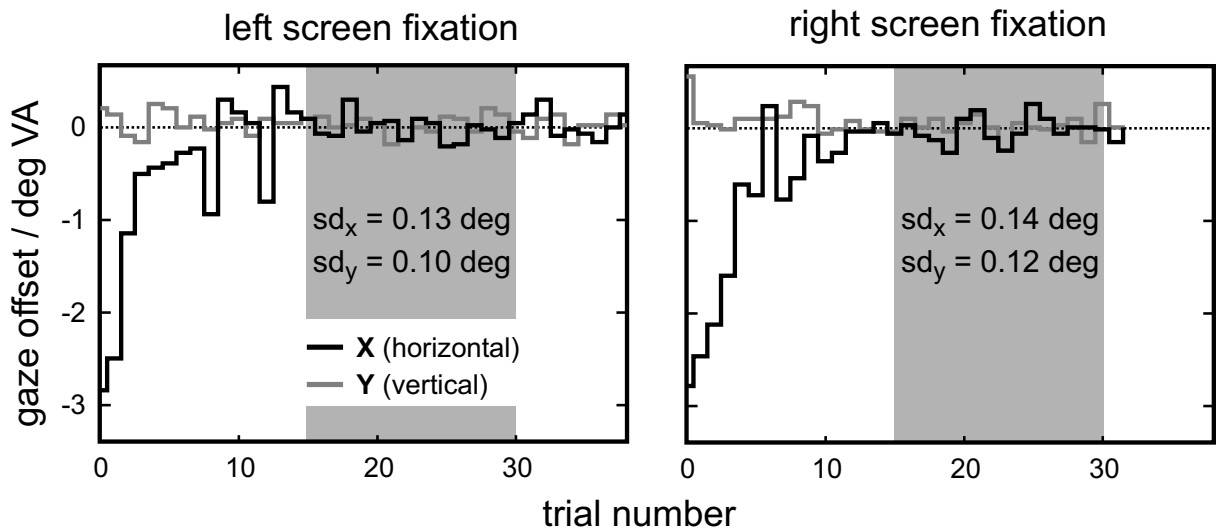
Figure 3.4 Convergence and residual uncertainty of stereoscopic pre-calibration procedure. Gaze offsets due to switching of the fixation spot to the lateral screens become successively smaller in the course of the pre-calibration procedure. Screens initially were misaligned with respect to horizontal but not vertical vergence in this example. With a feedback amplification of –1/3 about 10 trials are needed to correct an offset of 2.5° visual angle (VA). Beyond this the precision is not further improved because of restricted accuracy of the ocular fixation and gaze-direction determination, respectively. Residual uncertainty can be estimated from the variance during the stagnating phase (trials 15–30).

parameterization under the given conditions. We compared all available binocular CRFs determined two times with identical stimulation parameters in successive recording blocks (Fig. 3.6). The radial offset (euclidian distance) of CRF centers between two successive recordings is narrowly and left steeply distributed around a median of 0.025° (90% percentile = 0.070°, N = 195 from 15 sessions). A residual uncertainty of ±0.05° (3 arc min) is a conservative estimate for most recording sites.

3.3.3 Congruence of lateral screens

The positional calibration procedures are performed to optimize congruence of the lateral and central screens. To estimate the achievable precision of congruence we compared left and right eye to central CRF mappings, and left to right mappings, after a single run of the DCRF calibration procedure. Translation vectors between corresponding left and right eye and binocular CRFs are calculated to quantify congruence pairwisely for each of the 3 possible combinations of pathways left (L), binocular (B) and right (R): L–B, R–B, L–R. This means, for each recording session (day) three translation vectors result. Each vector is an average over the offsets of all channels revealing sufficient responsiveness via the two pathways in consideration. For the left-right congruence, for example, only those channels can be used to calculate the translation, that are sufficiently binocular, and therefore yield a quantifiable CRF
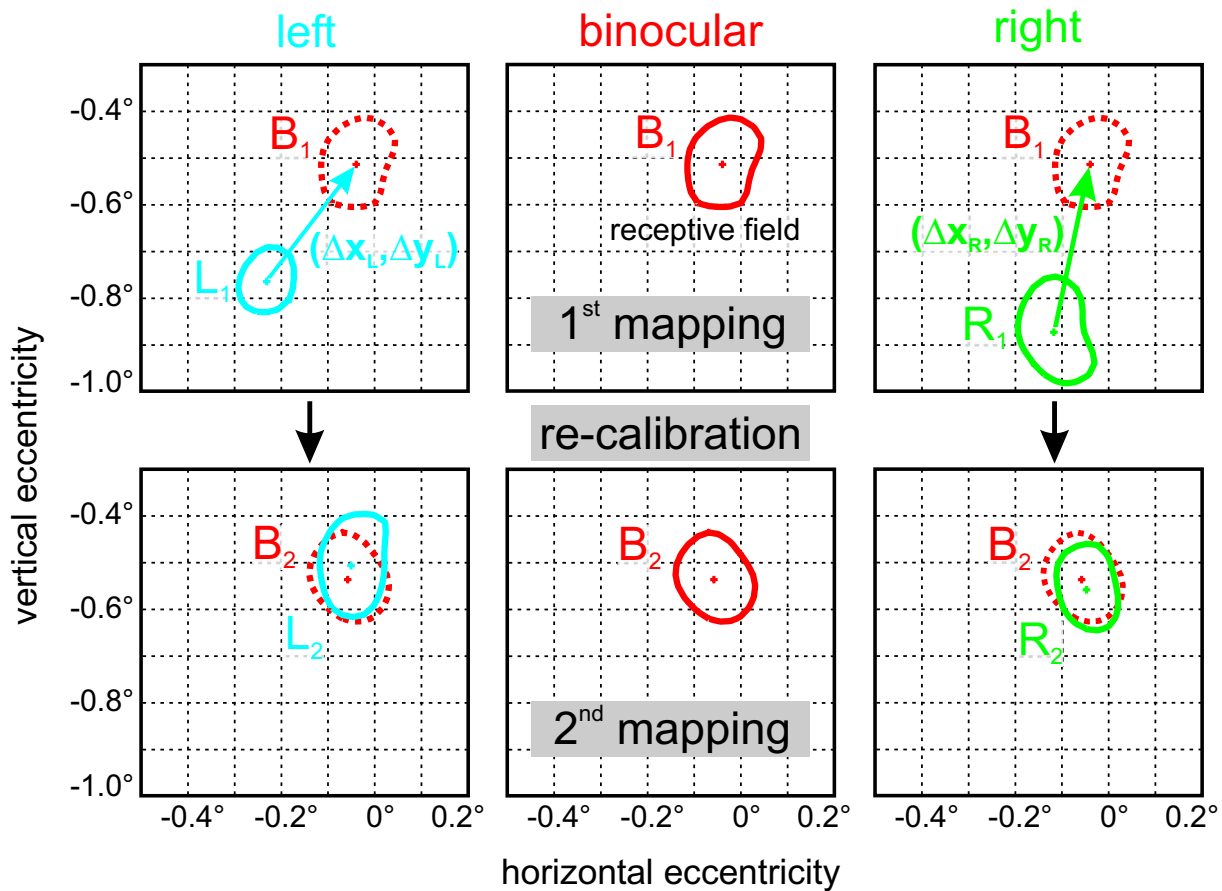
Figure 3.5 Principle of calibrating the stereoscope based on RF localization. Upper and lower row show RF contours (−3dB level) of a single recording site mapped two times in successive recording blocks with a re-calibration of the lateral screen coordinates in between. L, R, B denote the left and right monocular and binocular RF respectively. From the results of the first mapping (upper row) the incongruence of the left and right screen compared to the central one is judged from the offset of the respective RFs ($\Delta x_{L/R}, \Delta y_{L/R}$). (The dotted contours represent a copy of the binocular RF into the left and right maps for illustration purposes only.) To improve precision of this procedure the translation vectors for calibration of the left and right pathway are averaged over all RFs available for the respective side (only one channel shown). The success of the calibration is controlled in a second mapping (lower row). The remaining RF offsets (L_2–B_2, R_2–B_2, L_2–R_2) are used to estimate the precision of the calibration procedure (cf. Results and Fig. 3.7). The offsets of binocular RFs between two recording blocks ($B1$–$B2$) are used to estimate reliability of RF localization (cf. Results and Fig. 3.6) as the main restricting factor for overall precision of the method.

structure for left and right eye stimulation. The average number of evaluable CRFs for aligning two pathways was 10–11 (of maximal 16) for a typical recording session. Fig. 3.7 depicts the median absolute values of the translation vectors for the three pairs of pathways (N = 23 sessions). The congruences of the central screen to the left and right screen, respectively, are about the same (R–B: 0.023°, L–B: 0.019°), the congruence between left and right is worse (L–R: 0.032°). This precision is achieved with a single run of the DCRF-procedure, which on average consisted of 31 trials of 3692 ms mean recording time, i.e., 115 s (~2 min) total recording time, including about 1/2 min for recording the baseline condition (fixation without additional stimuli).

3.3.4 Ocular dominance

Beyond calibrating the setup, the DCRF method allows investigation of ocularly specific CRF parameters or interocular interaction. As a simple example evaluable for neuronal group activity, and with the same parameter settings as used for calibration, we present ocular dominance measurements. With the setup calibrated the DCRF-method inherently provides ocularity of the CRFs. Fig. 3.3D shows 16 left and right eye MUA CRF maps of a single session. Responsiveness of all channels shows a more or less pronounced selectivity for either eye. These selectivities are not randomly distributed in this example but are clustered with respect to eye dominance instead. Right eye dominance can preferably be found in a wedge-shaped area in the upper right corner (electrodes 3,4,7,8,10), while left eye dominance is prevalent in the other corners. Quantification of ocular selectivity with an ocular dominance index (ODI, cf. Methods) yields overall distributions as depicted in Fig. 3.8B. MUA is more ocularly selective (SD = 0.19, N = 392) than LFP (SD = 0.16, N = 351), but LFP is still sensitive to eye differences.
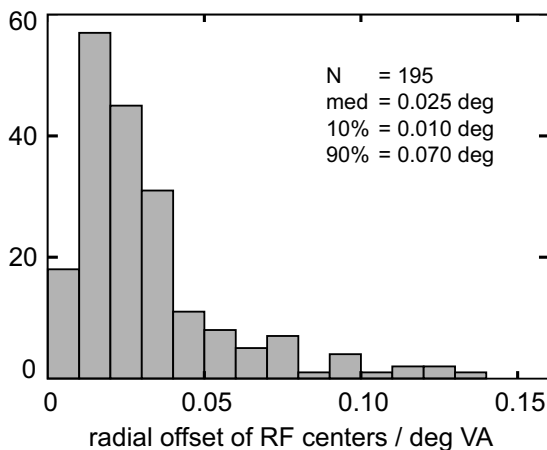
Figure 3.6 Reliability of RF localization. The histogram shows the frequency distribution of the positional displacement of all binocular RFs between two recording blocks within one session (cf. B_1–B_2 in Fig. 3.5). The median radial offset is 0.025° VA (10 % percentile: 0.01°, 90 %: 0.07°). The eccentricity of all 195 MUA RFs contributing to this distribution was ~0.3–1.8° VA.
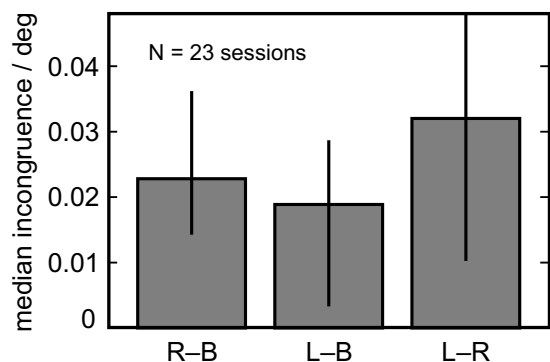
Figure 3.7 Precision of calibration procedure. The three bars denote the remaining incongruence between the left and right and central screens as estimated from the residual offsets among the left and right monocular (L, R) and the binocular (B) RFs. Bar height denotes the median value of 23 recording sessions (R–B: 0.023°, L–B: 0.019°, L–R: 0.032°; mean of these three medians: 0.025°). Error bars denote the 10 % and 90 % percentiles respectively (L–R 90 % end is clipped for depiction). For each day the screen offset was calculated from the average displacement of all RFs available for the pathways in consideration (cf. Fig. 3.5).

3.4 Discussion

We introduced a stereoscopic stimulation setup with natural eye vergence alignment and precise vergence–accomodation adjustment in restrained awake animals. The method used for calibration is at the same time suitable for investigating several aspects of binocular receptive field properties. In the present study we discuss limitations in spatial precision of the calibration procedure and demonstrate applicability and effectiveness of the dichoptical CRF mapping method with short-term recording ocularity measurements in awake macaque monkeys.

3.4.1 Calibration

The reliability of the single CRF localization is most critical for the precision of the calibration procedure as used here. Comparison of successive recording blocks revealed an inaccuracy of less than $0.05°$ (3 arc min) for most CRFs (median $0.025°$). This estimation is based on binocular CRFs mapped two times in consecution and we assume that it is transferable to the simultaneously mapped monocular CRFs. For the alignment of the lateral screens relative to the central screen the average offset of 10 to 11 simultaneously measured single CRFs was
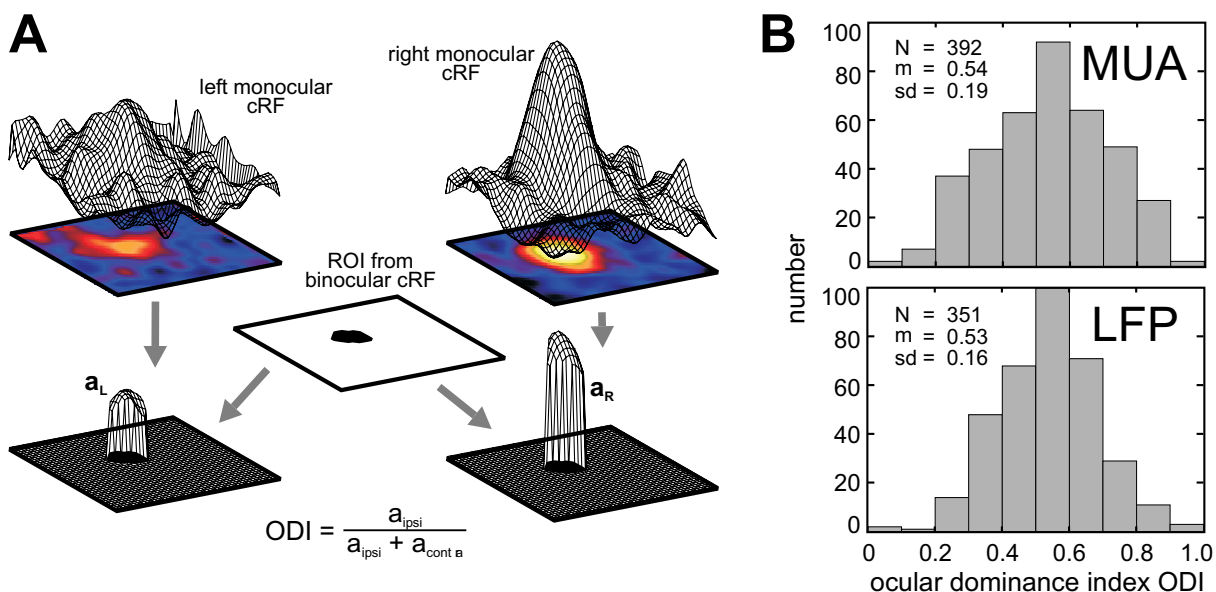
Figure 3.8 Ocular dominance measurements with the DCRF mapping. **A:** The left and right eye monocular RF maps can be compared to define an ocular dominance index ODI in terms of an ipsi- / contrallateral comparison: ODI = 0 contralateral response only, ODI = 1 ipsi only (Macy et al., 1982). The strengths of the respective left and right eye stimulus-response correlations ($a_{L/R}$) are calculated as average amplitude within a region of interest (ROI) defined by the –3 dB level of the binocular RF (cf. Methods). **B:** Distributions of ODI based on MUA and LFP data respectively. The more narrow distribution of LFP based ODI reflects the wider range of integration of this signal. But there is still a substantial specifity of LFP for ocularity (ODI = 0.25 (0.75) corresponds to a factor of 3 (1/3) in responsiveness difference between contra- and ipsilateral). All data is recorded in the supergranular layers 2/3 of V1 in two monkeys.

used in each session. This would improve the precision for the screen alignment, if the individual CRF offsets are independent in their size and direction. But the median incongruence between lateral screens and the central screen (Fig. 3.7) was of the same size as the individual CRF inaccuracy (Fig. 3.6). This means that for the calibration procedure no improvement was achieved by averaging over multiple CRF offsets within a single session. Closer inspection of the distributions of CRF offsets revealed that between two successive recording blocks at some days there was a coherent shift of all determined CRF centers in a similar direction. This could indicate that the inaccuracy of the CRF position was partly due to systematic errors rather than random fluctuations in CRF-measurements. But unexpectedly there is no correlation ($r < 0.2$; data not shown) between the precision of calibration and the relative or absolute size of the coherent CRF shifts. This means that the coherent shifts seem not responsible for the failure of precision improvement by averaging over multiple CRFs.

Alternatively, the assumption could be wrong that the reliability of the binocularly mapped CRFs is also valid for the monocularly mapped CRFs. Monocular CRFs of one eye are differently strong driven by the random stimulus, depending on the ocular dominance at the recording site. Accordingly, CRF localization becomes more difficult with weak drive and high maintained activities, leading to low signal-to-noise ratios. The singularity criterion (cf. Methods) at a single threshold level alone may not prevent mis-judgement of the CRF center when the peak in the spatial response profile is multi-modal due to a low signals-to-noise level. This uncertainty in CRF localization could probably be overcome by calculating CRF offsets with a 2-dimensional cross-correlation analysis among both monocular CRF maps. This avoids parameterization of the individual CRFs in consideration and might succeed even with weak CRFs when the peak detection algorithm for the individual CRF fails. The computational effort, on the other hand, is clearly higher, which is an argument especially in the case of (pseudo-)online evaluation. 2-D cross-correlation would also be suitable for evaluating offsets and similarity between left and right eye single unit CRFs with ON-OFF sub-structures. This did not play a role here since a) we recorded in the supragranular layers of V1, where complex-type cells dominate, and b) we concentrated on evaluation of CRFs based on population signals, namely MUA and LFP. In addition, with the low eccentricities used here (~0.2–2.0°) characterization of CRF sub-fields in awake animals would only be possible, if correction for the inevitable residual eye movements during fixation is possible (Livingstone and Tsao, 1999; but see Brinksmeyer et al., 2001).

Another limiting factor, which presumably restricts improvement of congruence, is the fact that the observed CRF uncertainties are in the order of magnitude of the used monitor resolution of the setup: a single pixel covered 0.022° visual angle. Such discrete stimulation limits the precision of the calibration procedure. Since in both monkeys CRFs were located parafoveally, we can not tell in which way the use of more peripheral CRFs would influence the precision of the procedure.

The worse congruence between left and right screen, compared to the congruence between left/right and central screen, is likely to result from the calibration procedure. The lateral screens were each aligned with respect to the central screen, but not with respect to each other. One has to expect a maximal error propagation of a factor of two. For applications which desire high congruence between left and right, direct alignment of the lateral screens with respect to each other could improve the situation. It can not fully be excluded, however, that the larger displacements between left and right monocular CRFs (compared to their displacements relative to the binocular CRFs) indicate some degree of disparity coding by positional shift even in MUA CRFs. Prince and collegues (2002b) did find a weak correlation between preferred disparities of SUA and MUA recordings from the same electrode in monkey V1, indicating some degree of clustering for disparity preference. They also found positional-coding (along with phase-coding) to be quite common for disparity-coding in monkey V1 (Prince et al., 2002a).

3.4.2 Applications

Beyond the calibration procedure, we demonstrated the applicability and effectiveness of the DCRF method by ocular dominance measurements. 100–120 s of total recording time were sufficient to map both monocular and binocular CRFs of 16 channels simultaneously. The time needed depends on the size of the visual area to be covered, in combination with the desired spatial resolution. Recording time grows linearly with the number of stimulus grid positions of one stimulation pathway. Simultaneous mapping of the additional pathways does not cost extra time because it is made in parallel. For our recordings we used a 4×4 electrode arrangement with 750 µm pitch in parafoveal V1. Correspondingly, CRF areas are largely overlapping, hence reducing the size of the scanning area. In addition, we were mainly interested in characterizing the CRFs with respect to their location and ocular dominance, and therefore used a spatial resolution of the sampling grid (16×16 covering all CRFs) which is at the lower end of the acceptable limit.

We found that the distribution of ocular dominance indices is wider for MUA than for LFP (Fig. 3.8B), which means that LFP is less selective for ocular dominance than MUA. Considering the larger effective range of integration (LFP: ~500 μm, MUA: ~50 μm) and the alternating organization of ocular dominance with a periodicity of ~1 mm/pair along the V1 / V2 border (LeVay et al., 1985) this is qualitatively consistent with expectation.

The DCRF method in principle is suitable for analysing mechanisms underlying binocular interaction. Once the setup is calibrated, single CRF areas can be sampled at higher spatial resolution by adjusting the stimulus grid. Due to the 3-screen design monocular and binocular CRFs are available at the same time. In a similar approach this has already be done with a standard stereoscopic setup in anaesthetized cats (Anzai et al. 1999b,c). Making use of 1-D oriented (bar-like) multi-focal m-sequence stimuli (Sutter, 1987), Anzai and collegues mapped 1-D monocular CRFs and binocular interaction RFs. The latter contain information on the non-linear interaction of left and right eye signals for different disparities orthogonal to the orientation of the oriented stimulus. In another approach Livingstone and Tsao (1999) mapped 1-D monocular CRFs and binocular interaction RFs in awake monkey with the use of sparse stimuli (2 single bars with variable disparity). To solve the eye vergence problem they used the technique of colour-coding the stimuli for left and right eye presented on a single screen, with appropriate filtering in front of each eye (anaglyphs), hence stimulating different colour channels in both eyes. Corresponding analyses on binocular interactions can also be conducted with our method. We use a sparse instead of a dense noise stimulus[1], which is more adequate here since the binocular stimulus is physically overlaid with the monocular stimuli. Our test stimulus can be spatially restricted to subfield size of a CRF (e.g., for simple cells), is not oriented, and of the same colour. This makes it suitable for investigating 2-D CRF interactions, i.e., effects of binocular stimulus disparity can be characterized in any direction irrespective of the preferred orientation of the neuron. Practically this could mean, for example, to look for the influence of stimulation with a test spot at various 2-D positions in one eye, while simulating at the CRF centre in the other eye. If the sequences of successive stimulus positions are statistically independent for the different pathways, and are sufficiently long in duration, all possible combinations of test spot positions will occur several times at equal probability. In contrast to our data presented here, one may include dark in addition to bright spot

---

1   The use of sparse versus dense noise has implications for the resulting CRFs even in 1st order (Reid et al., 1997), because the underlying system is nonlinear; for a systematic description of deterministic approaches in non-linear system analysis see Sutter (1992).

stimulation to improve distinction of sub-regions when investigating CRF structures (e.g., Eckhorn et al., 1993b).

One problem, inherent to stereoscopic setups, is especially critical for investigations of binocular interaction: mis-aligned eye vergence. It can only provisionally be overcome by reference-cell methods (Hubel and Wiesel, 1970; Ferster, 1981; Anzai et al., 1999a). With our improved stereoscope and advanced calibration method this problem is reduced to a minimum. The estimated precision in screen congruence of 0.025° visual angle is small, compared to the range of preferred disparities found in V1 neurons (Anzai et al. (1999a) in cat: SD = 0.68°/0.44°, phase and position disparities, respectively; Prince et al. (2002a) in monkey: SD = 0.26°/0.21°), and compared to the flankwidth of disparity tuning curves (0.2–0.5°; cf. examples in Poggio et al., 1988; Prince et al., 2002a,b). Our setup offers an additional control option to avoid the mis-alignment problem: Genuine binocular CRFs (mapped via the binocular pathway) can be compared with the binocular interaction CRF at zero disparity (mapped dichoptically with stimuli at corresponding positions between left and right). Both CRF positions should be identical when the setup is calibrated correctly.

Our setup may also be useful for other applications. In binocular rivalry paradigms incongruent dichoptical stimulation easily leads to diplopia, unless strong additional congruence forcing visual stimuli are presented. When the congruent part of the visual stimulus only consist of the fixation spot diplopia is less probable in the case of a binocularly viewed spot (central screen) compared to a dichoptically congruent pair of spots on the lateral screens (personal observation). In studies on visual object coding one could superimpose the sparse density test stimulus for CRF mapping via the lateral screens, while simultaneously presenting a behaviourally relevant visual object on the central screen. This allows investigations of modifications in CRF properies due to visual stimulation or, for example, due to attentional effects.

3.4.3 Conclusion

Our enhanced stereoscopic setup together with the DCRF calibration method is a useful tool for dichoptical stimulation in awake restrained animals. The possibility of simultaneous binocular and dichoptical stimulation with high precision in eye vergence alignment and low vergence-accomodation mis-match combines the advantages of classical stereoscopes with those of shutter- or anaglyph-systems, and makes it especially suitable for investigating the neuronal basis of stereovision.

3.5 Appendix – Receptive field cinematogram

For each of the 256 potential positions of the 16×16 hexagonal grid the stimulus function $S_{nm}(t)$ is a time course indicating spot presence at the grid node (n, m) with a value of 1, and spot absence with a value of 0. The temporal luminance profile $P(t)$ of a single spot presentation is

$$P(t) \;=\; \text{rect}\left(\frac{t - D/2}{D}\right) \;=\; \begin{cases} 1: & 0 \leq t < D \\ 0: & \text{else} \end{cases} \quad . \tag{A3.1}$$

The time course of stimulus onsets $T_{nm}(t)$ at each grid node (n, m) can be described as series of dirac delta functions

$$T_{nm}(t) \;=\; \sum_{i=1}^{N_{nm}} \delta(t_i^{nm} - t). \tag{A3.2}$$

Note, the stimulus position is a function of time, i.e., at each time t_i^{nm} the spot is present at only one grid node (n, m). The stimulus function can be written as a convolution

$$\begin{aligned} S_{nm}(t) &\;=\; \int P(\tau) T_{nm}(t - \tau)\, d\tau \\ &\;=\; \sum_{i=1}^{N_{nm}} P(t - t_i^{nm}) \;\;, \end{aligned} \tag{A3.3}$$

which reduces to a sum making use of the property of the dirac function. N_{nm} is the number of spot occurrences at each grid node within the recording epoch T_R.

The same can be done for the space dimensions. The spatial luminance profile $L(x, y)$ of a single spot is of gaussian shape with maximum intensity I_0

$$L(x,y) \;=\; I_0 \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \tag{A3.4}$$

The stimulus grid can be described as a sum of spatial dirac functions

$$g(x,y) \;=\; \sum_{n,m} \delta(x_n - x, \, y_m - y) \tag{A3.5}$$

with

$$\begin{aligned} x_n &\;=\; x_0 + \left[(n + (-1)^m/2)\, a\right] \\ y_m &\;=\; y_0 + \left[\sqrt{3}/2 \; ma\right] \end{aligned} \qquad n, m = 1, 2, \ldots R \;\;, \tag{A3.6}$$

where (x_0, y_0) are the coordinates of the lower left corner of the simulus grid, a denotes the grid constant (scaling the to be scanned area in size), and R denotes the resolution which is 16 in our case. The stimulus function can then be generalized to

$$
\begin{aligned}
S(x,y,t) &= \iint S_{nm}(t) L(\xi,\upsilon) g(x-\xi, y-\upsilon) \, \mathrm{d}\xi \, \mathrm{d}\upsilon \\
&= \sum_{n,m} \sum_{i=1}^{N_{n,m}} P(t-t_i^{nm}) L(x-x_n, y-y_m) ,
\end{aligned}
\tag{A3.7}
$$

again making use of the property of the dirac function. This represents a continuous description of the stimulus function, i.e., the actual physical stimulus, apart from the luminance decay function of the screen phosphor.

For the purpose of this study we did not consider the real physical luminance time course for each position in space. Instead the spatio-temporal RF profiles are determined by cross-correlating the mean-free neuronal response of each electrode $R^e(t)$ with the stimulus onset time courses $T_{nm}(t)$ at the grid node (n, m)

$$
\begin{aligned}
RF_{nm}^e(\tau) &= 1/N_{nm} \int_0^{T_s} R^e(t+\tau) T_{nm}(t) \, \mathrm{d}t \\
&= 1/N_{nm} \sum_{i=1}^{N_{nm}} R^e(t_i^{nm}+\tau).
\end{aligned}
\tag{A3.8}
$$

The correlation is normalized to the number of stimulus presentations N_{nm} at each grid node, not to the signal power of the neuronal response. The spatial and temporal profile of the test spot itself are ignored in this way of stimulus-response correlation. Therefore, the RF functions are equal to peri-stimulus time histograms (PSTHs) calculated for stimulation at each individual grid node.

For the real physical stimulus $S(x,y,t)$ the luminance time courses at the individual grid nodes are eventually not completely statistical independent any more, depending on the ratio between the width of the spot's luminance distribution (σ in A3.4) and the grid constant a (typical values $[\sigma, a]$ in arc min: monkey H $[4.13', 6.88']$, S $[2.75', 4.13']$). Both, the dwell time of individual spot presentations and the size of the gaussian patch are adjusted to optimize MUA response properties. With respect to a system-theoretical white noise approach (Marmarelis and Marmarelis, 1978) they are not adapted to the transfer properties of the visual system, in the sense that the stimuli are more band-limited in time and space than the visual system. The resulting RF structures therefore are stimulus specific rather than an ideal system characterization independent of the specific stimulus function. We accepted this compromise,

since for our purpose we were merely interested in determining the MUA RF localizations of all recording sites in the shortest possible recording (and evaluation) time. The center of mass of an RF should essentially not be influenced by a slight temporal undersampling or spatial blur of the RF structure. In contrast, RF sizes and the correlation length of the neural response to the stimulus might be overestimated with the parameter settings in use. Nevertheless, for calculation of the RF centers with sub-grid resolution, RF maps were spatially interpolated to original CRT pixel resolution by low-pass filtering according to the sampling theorem.

4 Binocular Rivalry

Perception-related modulations of local field potential power and coherence in primary visual cortex of awake monkey during binocular rivalry

Cortical synchronization at γ-frequencies (35–90 Hz) has been proposed to define the connectedness among the local parts of a perceived visual object. This hypothesis is still under debate. We tested it under conditions of binocular rivalry (BR), where a monkey perceived alternations among conflicting gratings (objects) presented singly to each eye at orthogonal orientations. We made multiple microelectrode recordings of multi-unit spike rate (MUA amplitude) and local field potentials (LFP) from striate cortex (V1) during BR while the monkey indicated his perception by pushing a lever. We analysed spectral power and coherence of MUA and LFP over 4–90 Hz. As in previous work, high coherence was found when pairs of recording locations were stimulated congruently in both eyes at their preferred orientation. With incongruent (rivalrous) stimulation LFP power was often modulated in consonance with the perceptual state. This was not visible in MUA. The perception-related LFP-modulations occurred at low- and medium-, but not at γ-frequencies. Perception-related modulations of LFP coherence were also restricted to the low–medium range. In conclusion, our results do not support the expectation that γ-synchronization in V1 is related to the perceptual state during binocular rivalry, but instead suggest a perception-related role of synchrony at low and medium frequencies.

4.1 Introduction

Synchronization can occur in the visual cortex at γ-frequencies (35–90 Hz) among neurons activated by a visual object. This synchrony has been assumed to label the distributed features of

the object for belonging perceptually together. However, this hypothesis is still under debate. Tests, whether perception of a visual object is really paralleled by γ-synchrony may well be analysed in tasks where the visual stimulus is identical while the percept has two alternatives. This is the case with binocular rivalry (**BR**). BR describes the alternating perception of two incongruent visual stimulus patterns presented dichoptically to both eyes.

Binocular rivalry. There are ongoing controversies at different intermingled levels of description about the nature of binocular rivalry. From a long tradition of psychophysical experiments a debate arose whether binocular rivalry is caused by interocular competition (Lehky, 1988; Blake, 1989; Lee and Blake, 1999) or competition of coherent percepts (Diaz-Caneja (1928) translated in Alais et al., 2000 and replicated in Ngo et al., 2000; Kovacs et al., 1996, Logothetis et al., 1996). At the neurophysiological level this reads as competition between monocular channels at early stages (Blake, 1989) on the one hand, and competition of high level cortical representations (Logothetis, 1998) on the other hand. The strongly increasing relative number of single cells along the ventral pathway, that reflect the monkey's perception in binocular rivalry tasks, is taken as evidence for the high level explanation (Leopold and Logothetis, 1996; Sheinberg and Logothetis, 1997; Logothetis, 1998). Modulation of monocular channels in primary visual cortex due to alternating perception has been reported with fMRI in human (Tong and Engel, 2001), supporting the low level model.

Visual awareness. Besides identifying the locations and mechanisms that underly binocular rivalry, respective results are also discussed in the context of visual awareness and its neuronal correlates. Especially the role of primary visual cortex in containing signals correlated with visual awareness is highly controversial (Logothetis, 1998; Engel et al. 1999; Tong and Engel, 2001; Andrews, 2001). More recently the understanding gains acceptance, that computational resources along the whole visual pathway are involved in the task of resolving perceptual ambiguities, binocular rivalry being only one among others, like bistable figures of any kind (Blake and Logothetis, 2002). Not only the locus but also the quality of a potential awareness-relevant neuronal signal remains unclear. For primary visual cortex alone argumentation is based on such different measures as mean spike rate of single cells in awake monkey (Leopold and Logothetis, 1996), multi-unit spike correlation and spike-to-field-potential coherence in strabismic cat (Fries et al., 1997), and BOLD mass signal in fMRI of the human blind spot region (Tong and Engel, 2001).

Binding-by-synchronization. Synchronized γ-oscillations observed in primary visual area of cat (Eckhorn et al., 1988; Gray et al., 1989) and monkey (Kreiter and Singer, 1992; Eckhorn et al., 1993a) have often been associated with feature binding (reviewed in Eckhorn, 1999; Engel et al., 1999; Gray, 1999), and therefore should also play a role in perceptual segregation and integration of those features belonging to the currently perceived object. In the context of binocular rivalry the synchronization hypothesis has so far only been investigated intra-cranially in preliminary studies in awake monkey in our group (Kottman et al., 1996) and in strabismic cats (Fries et al., 1997), both suggesting a key role of γ-synchronization in awareness-relevant signalling.

As the hypothesis of object integration by synchronization at γ-frequencies is still under debate, we tested it in the present study under conditions of BR, where a monkey perceived alternations among conflicting gratings objects presented dichoptically at orthogonal orientations. The two alternative percepts occur despite identical physical stimulation. This allows to dissociate modulations of the neuronal activity, e.g., synchronization in the γ-range or at other frequencies, due to the perceptual state from those due to changing stimulation. For this we performed multi-channel microelectrode recordings from monkey primary visual cortex (V1) during binocular rivalry and analysed multi-unit activity and local field potentials over a broad frequency range to figure out the role of synchronization in V1 for conscious perception (visual awareness) during binocular rivalry.

4.2 Methods

4.2.1 Visual stimulation and behavioural task

Visual stimulation setup. Dichoptical visual stimulation was realized with a 3-screen setup suitable for humans and monkeys (Fig. 4.1A). Each monitor covered $18.2 \times 13.6°$ of the visual field with a resolution of 800×600 pixel driven at $98\,\text{Hz}$ frame rate. All 3 screens were synchronized and made congruent in view by means of two semi-transparent mirrors. A two step calibration procedure ensured adjustment of the screens with an accuracy of about $0.025°$ when using the setup with monkeys (Gail et al., 2002).

Binocular rivalry task. To induce binocular rivalry we used stationary, soft-edge, high-contrast, sinusoidal luminance grating patches of horizontal and vertical orientation, respectively
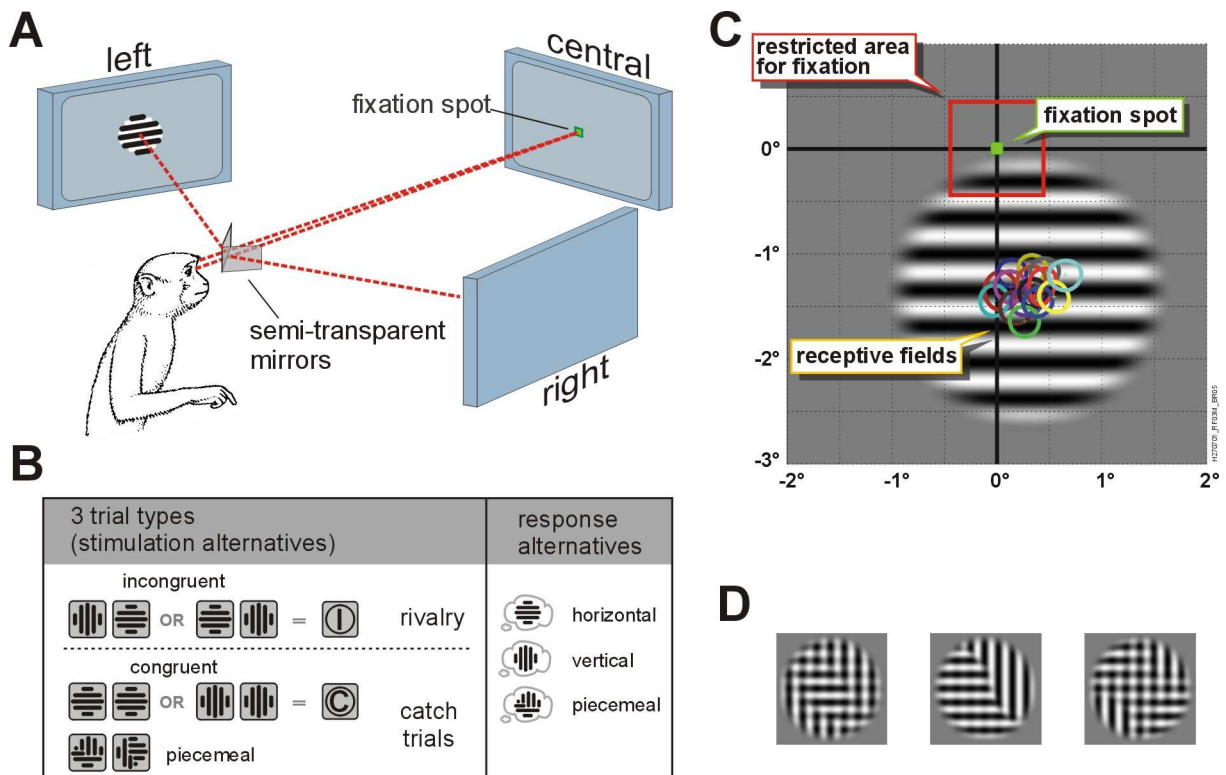
Figure 4.1 Visual stimulation setup and task. **A:** Dichoptical stimulation via semi-transparent mirrors. All screens are synchronized at 98 Hz (800 × 600) and precisely congruent in view. The central screen is viewed binocularly and is used for fixation with natural vergence during dichoptical tasks and for standard binocular RF mapping stimulation. Grating stimuli in the rivalry task are presented via the left/right screen. **B:** 3-alternative decision task. The paired pictograms for the different trial types denote the left and right stimulus, respectively. Incongruent left/right stimulation (Ⓘ) with orthogonal gratings causes binocular rivalry and is applied with variable relative contrast between left and right to control the monkey's psychometric performance. 2 types of catch trials are designed to evoke a predictable percept and serve as an additional control (details cf. Methods). **C:** Grating stimuli and receptive fields. Stimuli have full contrast (luminance via mirror: 0.5–54 cd/m^2; spatial frequency: 2.9 cyc/°) over 75 % of their diameter and a cos^2-shaped soft edge. The patch size (monkey H: 2.75°; monkey S and human: 1.83° diameter) and position was adjusted to cover all receptive fields of a single recording session within the patch's radius of full contrast (original scale example; RF contours denote -3dB amplitude). **D:** Examples of *piecemeal* stimuli. Two different of these mixed grating stimuli are presented dichoptically in the piecemeal catch trials and can not be fused to a percept of uniform orientation. They mimic the percept when exclusive dominance of one of the grating stimuli in rivalry trials fails.

(Fig. 4.1). They were dichoptically presented at corresponding retinal positions of the left and right eye while the monkey had to keep fixation within ±0.45° at a small spot on the central screen. The patch size and position was adjusted to cover with its full-contrast range the classical receptive fields of all recording locations in a single session. Patches were made as small as possible to maximize the probability of exclusive dominance of one of the gratings across its entire extension (O'Shea et al., 1997).

To control the monkey's behaviour we used a high portion of catch trials and permanently compared the psychometric performance with that of human observers recorded in this setup. Three different types of trials occurred in pseudo-random order: During *incongruent*

trials (~75 %) orthogonal gratings were presented to the left and right eye, reliably evoking binocular rivalry. *Congruent* trials (~15 %) consisted of gratings at identical orientations, which always become perceptually fused to a uniform grating. *Piecemeal* trials (~10 %) consisted of two different patches applied to the left and right eye, where each was a mixture of horizontal and vertical grating components (Fig. 4.1D). The left and right patch were not complementary and therefore in no way could be fused to the percept of a grating with uniform orientation, i.e., this always induced a *piecemeal* percept. Irrespective of the trial type, the monkey had to indicate whether he perceived a horizontal (lever up), a vertical (down) or a piecemeal patch (lever release). The trial ended after this decision had been made.

When rivalling stimuli are switched on, exclusive dominance of one of the two stimuli is not immediately established, but the two incongruent stimuli appear somehow fused instead (Wolfe, 1983). We took this into account in two ways. Firstly, we inserted a piecemeal epoch of random duration (50–250 ms) at the beginning of the congruent trials (Fig. 4.2B). This made the congruent trials look more similar to what the incongruent trials were expected to look like. Besides this, the monkey was thereby teached not to report seeing piecemeal immediately, but to wait whether the stimulus turnes out to become uniform in orientation or not. The second measure further reinforced this behaviour. Reports of seeing piecemeal were not accepted before 1200 ms post-stimulus onset. Expiry of this interval was indicated to the monkey by dimming of the fixation spot. Conversely, reports of horizontal or vertical orientation were not accepted after this point in time (Fig. 4.2). Taken together, the monkey's task was in any case to start a trial by touching the lever, to wait for the stimulus to appear, and then to wait for the stimulus, being initially piecemeal, to turn into a grating of uniform orientation

Figure 4.2 Trial timing. Depending on the trial type different behavioural responses in distinct epochs were considered correct and reinforced by reward. **A:** In rivalry trials (Ⓘ: incongruent gratings from the beginning on) dominance of one grating orientation can be reported until dimming of the fixation spot at 1200 ms post stimulus onset (dark blue epoch). Failure of dominance of one of the two stimuli within this time (i.e., enduring piecemeal percept) can be reported afterwards (light blue). **B:** Catch trials with congruent stimuli (Ⓒ) only allow report of an explicit orientation (dark blue epoch) after the initial piecemeal stimulation of variable duration (black). **C:** Catch trials with enduring piecemeal stimulation (black) only allow piecemeal report after dimming of the fixation spot (light blue).
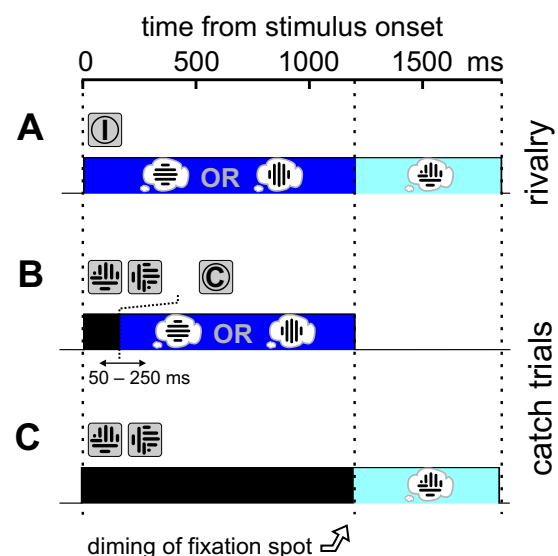
before the fixation spot dimmed. If this was the case he had to indicate the orientation of the grating by a corresponding lever switch, if not he had to release the lever after dimming of the fixation spot.

The congruent and the piecemeal trials served as catch trials: Perception was predictable in these cases and only one response alternative was correct and rewarded. They mimicked two alternative ways of experiencing the incongruent trials, where perception was unpredictable by the experimenter and both response schemata (early switch or late release) were allowed and rewarded.

We took an additional precaution to control the monkey's behaviour. In the incongruent trials we pseudo-randomly varied the contrast difference between the left and the right grating from trial to trial. The probability of perceiving, for example, the horizontal orientation should then depend on the contrast difference between horizontal and vertical grating in a way comparable to the respective human psychometric function collected with the very same setup. Difference in relative contrast ranged from –100% (left grating only) to +100% (right grating only) in 7 (monkey H) and 9 (monkey S) steps, respectively. Those trials with monocular stimulation and hence non-ambiguous perception served as yet another sort of catch trials.

4.2.2 Data aquisition and analysis

Animal preparation. Two male rhesus monkeys (Macaca mulatta) aged 9 (monkey H) and 14 years (monkey S) participated in the experiments. All procedures were carried out in accordance with German laws of animal maintenance and experimentation and the guidelines published in the NIH *Guide for the Care and Use of Laboratory Animals* (NIH publication no. 86–23, revised 1987). After intensive training and shortly before the experimental sessions a plastic chamber (10 mm o.d.) was implanted under deep barbiturate anaesthesia to give access to V1 through the intact dura. In both monkeys the chamber was implanted close to the lunate sulcus to obtain parafoveal receptive fields. 3 stainless steel headposts ensured painless head fixation during recording sessions. They had been implanted years before.

Recording. In each session up to 16 quartz-isolated, platinum-tungsten fiber-microelectrodes were individually advanced into the cortex under acustical and optical control of the recorded signals (Eckhorn and Thomas, 1993). After detecting the first reliable spikes they were slowly driven another 150–250 μm. Daily cleaning the dura before and more intensively after each

session reduced thickening of the dura (and hence dimpling during electrode insertion) to a minimum, ensuring recordings from layers 2 and 3. The electrodes were arranged in a regular 4-by-4 array (750 µm pitch). From each raw broad-band signal (1 Hz to 10 kHz) we separated multiple unit activity (**MUA**) by band-passing (1–10 kHz; 18 db/oct), full-wave rectifying and subsequent low-pass filtering (140 Hz; 18 db/oct), yielding an amplitude-weighted measure of population spike density near the electrode tip without rejecting low amplitude spikes. The mean MUA amplitude during prestimulus recording (blank grey screen with fixation spot) was subtracted from the following response epochs. Second, local field potentials (**LFP**) were obtained by band-passing from (1–120 Hz, 18 db/oct). Both analog signals (MUA, LFP) were sampled at a rate of 500 Hz.

Data analysis. Sliding-windows (128 ms epoch length; 32 ms shifts) were used to calculate time-resolved spectral power at single site recordings and coherence between signals at pairs of sites. For the sake of direct comparison the same windows were used for the single-channel analyses of MUA amplitude. Spectra were calculated via Fast Fourier Transform after applying a Hamming window to the mean-free signal of each epoch. Pairwise coherence between signals of two different electrodes n and m were calculated using Bartlett-smoothing across N trials with identical conditions (Glaser and Rushkin, 1976):

$$\hat{\kappa}_{nm}^{2} = \frac{\left| \sum_{i=1}^{N} S_{n,i}(f,t) \ S^{*}_{m,i}(f,t) \right|^{2}}{\sum_{i=1}^{N} |S_{n,i}(f,t)|^{2} \ \sum_{i=1}^{N} |S_{m,i}(f,t)|^{2}} \ , \tag{4.1}$$

where S is the complex Fourier spectrum of the signal epoch centered around time t (asterisk denotes the complex conjugate). The expected bias, i.e., the random coherence of the estimate depending on the number of trials, was subtracted (Benignus, 1969; in: Glaser and Rushkin, 1976):

$$\kappa^{2} = \hat{\kappa}^{2} - \frac{1}{N}(1 - \hat{\kappa}^{2}). \tag{4.2}$$

In short, coherence is a sensitive measure to estimate the linear correlation between two signals independently at each frequency. To get a high coherence value at a given frequency a high covariation of the spectral amplitude and a constant phase difference (not necessarily zero) at this frequency across trials is needed. For each measure signal epochs were either aligned with respect to stimulus onset or to the point in time of the monkey's behavioural re-

sponse to average data across trials. We divided the spectrum in three sections approximating the classical theta/alpha-, beta- and gamma-ranges : *low* (4–12 Hz), *medium* (12–27 Hz) and *high* (28–90 Hz).

Differences in spectral power between two conditions were tested by means of a Student t-test ($\alpha = 0.05$), where the number of trials defines the sample size. Coherence analysis for each frequency bears only a single value per condition, per time epoch and per electrode pair. Tests based on the comparison of estimated distribution parameters (e.g., mean in t-test) are a priori not possible. Therefore, confidence intervals were estimated, and the difference of two coherence values was judged by means of the overlap of their 95 % confidence intervals. To calculate confidence intervals, firstly the variance was estimated from the coherence value and the number N of epochs contributing (Glaser and Rushkin, 1976):

$$\sigma^2 \;=\; \frac{1}{2\,(N-1)} \left(1 - 0.004^{\,1.6\left(\frac{N}{N+1}\left(\kappa^2 + \frac{1}{N}\right) + 0.22\right)} \right) . \tag{4.3}$$

The 95 % confidence interval is then given by

$$FZT^{-1}(\kappa - 2\sigma)^2 \;<\; K^2 \;<\; FZT(\kappa + 2\sigma)^2, \tag{4.4}$$

were FZT is the Fisher's Z transform $FZT(x) = \frac{1}{2}\ln((1+x)/(1-x))$. Coherences were considered different, if their confidence bands did overlap less than 30 %. This threshold percentage was empirically determined; it corresponds to the percental overlap of the 95 % confidence intervals of two normal distribution when shifted just so far that a t-test becomes significant at the 0.05 % niveau.

Receptive field characterization. The binocular, and left and right eye monocular classical receptive field (**CRF**) positions were determined simultaneously with a newly developed dichoptical mapping technique (Gail et al., 2002) based on a unifocal reverse-correlation method (Eckhorn et al., 1993b). Applied with independent stimulus sequences simultaneously on all three screens it provides position, temporal dynamics, and ocular dominance of all CRFs of a session within 100–200 s recording time. Pseudo-online evaluation of these data allowed precise calibration of the dichoptic setup and adjustment of the visual stimulus. CRFs were lying parafoveally and close to or around the vertical meridian in both monkeys. Horizontal eccentricity: 0.4° (contra) to 0.2° (ipsi) in monkey S, 0.7° (contra) to 0.2° (ipsi) in monkey H. Vertical eccentricity (lower hemifield): 0.2° − 0.7° (S), 0.9° − 1.8° (H).

Ocular dominance was determined from the dichoptic mapping (Gail et al., 2002) and quantified by an ocular dominance index (**ODI**; Macy et al., 1982)

$$ODI = \frac{a_{ipsi}}{a_{ipsi} + a_{contra}} \quad, \tag{4.5}$$

where $a_{R/L}$ denotes the neural response by means of MUA/LFP amplitude to right/left stimulation within the receptive field. A monocularity index (**MI**) was defined to quantify deviation from the binocular equilibrium (LeVay and Voigt, 1988):

$$MI = 2|ODI - 0.5|. \tag{4.6}$$

This index ranges from 0 (equal responses for both eyes) to 1 (response only for one eye).

Perception-related modulations. One important rationale of our study is to decide whether , and if so what, aspects of V1 activity are correlated with perception. The procedure was to find recording sites that become modulated by different stimulus orientations in the congruent condition with respect to a certain measure, e.g., spike rate or LFP power. Unable to decide whether this modulation is due to the different visual stimuli or due to the accompanying different percepts, we compared it to the modulation of the same measure by different percepts in the incongruent condition, where the percept (but not the stimulus) changed from trial to trial. Note that for this analysis only trials with *no difference in stimulus contrast* between left and right were evaluated. In case that the modulation was present consistently in both conditions, we call it *perception-related*. In case that is only present in the congruent condition, we call it *stimulus-related*. Significance of the differences between vertical- and horizontal-report trials is tested separately in each condition (congruent and incongruent) and for each frequency bin (7.8 Hz resolution). Within the medium- and high-frequency ranges (cf. above) the alpha criterion is conservatively corrected for multiple testing (Bonferoni correction) by dividing by the number of frequency bins within this range (low range = 1 bin). It is then looked whether there is a significant difference anywhere within this range. For modulations to be considered perception-related, they have to show up in both conditions (congruent and incongruent), in the same direction (e.g., stronger for horizontal stimulus and percept), and at the same frequency (not only the same frequency range). Modulations (**M**) for each condition are quantified by an index

$$M = \frac{(R_{pref} - R_{nonpref})}{(R_{pref} + R_{nonpref})}, \tag{4.7}$$

where $R_{pref/nonpref}$ represents the response to the preferred/nonpreferred stimulus orientation (determined in the congruent condition). The response is either measured by spike rate (MUA amplitude) or by MUA/LFP power or coherence in a certain frequency range.

4.3 Results

4.3.1 Psychophysics

Four observations indicate that the monkeys understood and correctly performed the task. 1) The monkeys reported correctly in the catch trials. General performance, i.e., the percentage of trials with appropriate fixation and timing, was in both monkeys around 90 % (S: 85.6 %, H: 90.6 %). During ambiguous stimulation (left–right incongruence and contrast difference $<100\%$) we considered all decisions as valid that were made within the appropriate time interval. During non-ambiguous stimulation (left–right congruence or monocular) decisions additionally had to fit the stimulus. The average percentage of correct decicions from all valid non-ambiguous catch trials was 94.2 % (S: N=9 sessions) and 98.9 % (H: N=11), respectively. 2) The monkeys' psychometric function is nearly identical to those of humans (Fig. 4.3 A). The probability of perceiving the one stimulus or the other, depending on the contrast difference between them, shows in both monkeys the same sigmoidal run of the curve as the average curve of seven human observers using the identical stimulation setup. 3) Reaction delays are longer around zero contrast difference. This is the case in both monkeys and human (Fig. 4.3C): responses were 167 ms (S) and 98 ms (H), respectively, slower with incongruent stimuli at zero contrast difference compared to the average reaction time in the monocular or congruent conditions. 4) The probability of piecemeal reports peakes at zero contrast difference. This comes up to expectation equivalently to the longer reaction delays. Maximum rates of piecemeal reports in the incongruent condition with othogonal gratings (not piecemeal catch trials) were 5.0 % (S) and 11.2 % (H), respectively (Fig. 4.3B). Note that for human observers the average reported piecemeal probability was much higher (up to 50 %), but with large inter-individual differences.
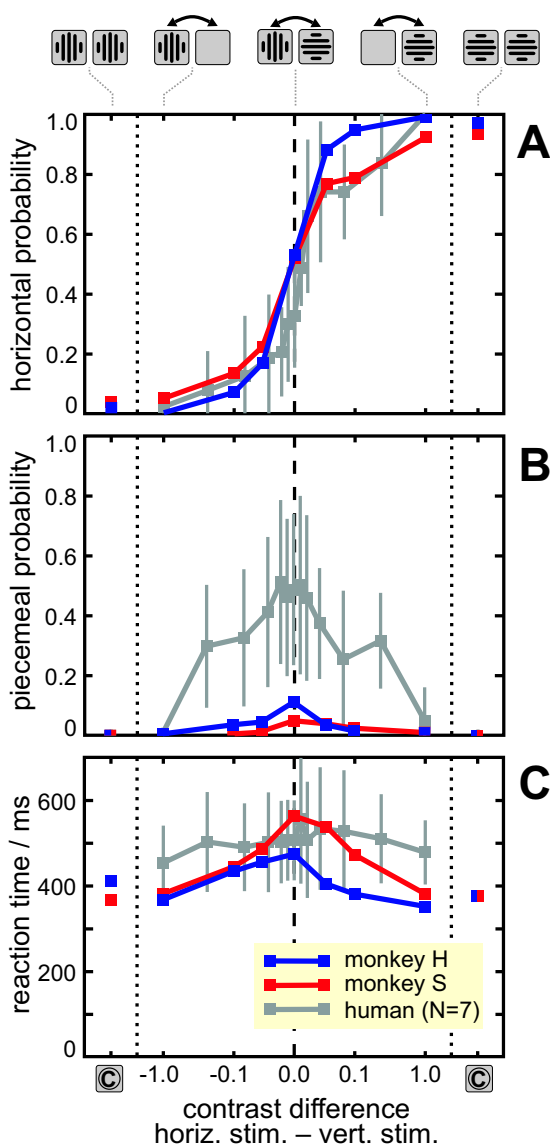
Figure 4.3 Psychometric performance. Comparison of mean human (N = 7, error bars: inter-individual variance) and individual overall monkey performance, depending on the difference in contrast between the horizontal and the vertical stimulus, reveals high similarity. Note the non-linear x-scale to enlarge the range of small contrast differences in the middle, and the non-continuity to add the two congruent conditions (ⓒ, monkey data only) at the very left and right, respectively. **A:** The probability of reporting the percept of a horizontal grating shows a typical sigmoidal dependence. At equal left/right contrast ($x = 0$) both monkeys show a 50/50-probability to report perception of the horizontal/vertical stimulus. **B:** The probability of reporting piecemeal perception in both monkeys is very low compared to human data, but shows a peak at zero contrast difference, in accordance with expectation. Human data show a high inter-individual variance. **C:** Mean reaction times also peak at zero contrast difference. Reaction times in the congruent conditions are measured with respect to the time of the stimulus switch, i.e., when the stimuli become congruent after the initial piecemeal epoch. All three findings together strongly suggest, that the monkeys experienced and honestly reported binocular rivalry at equal left/right contrast. Only these trials with equal contrast, together with the congruent trials as reference condition, were used for evaluation of perception-related modulations.

4.3.2 Data selection

We recorded data at 135 (monkey S) and 165 (H) recording sites during 9 (S) and 11 (H) sessions with a sufficient number of trials and reliable psychophysical performance. Only sites without technical artifacts and with clear MUA receptive fields were further analysed (S: 119, H: 160). For calculations of inter-electrode coherence remained 734 (S) and 1086 (H) electrode pairs, respectively. For the selected channels any individual trial was rejected containing a signal artifact in LFP according to visual inspection.

4.3.3 Congruent versus incongruent stimulation

The average spectro-temporal properties of the V1 activities are different between the two monkeys and with congruent compared to incongruent stimulation (Fig. 4.4). LFPs of monkey H show a prominent stimulus-induced power increase between 40 and 60 Hz in the congruent
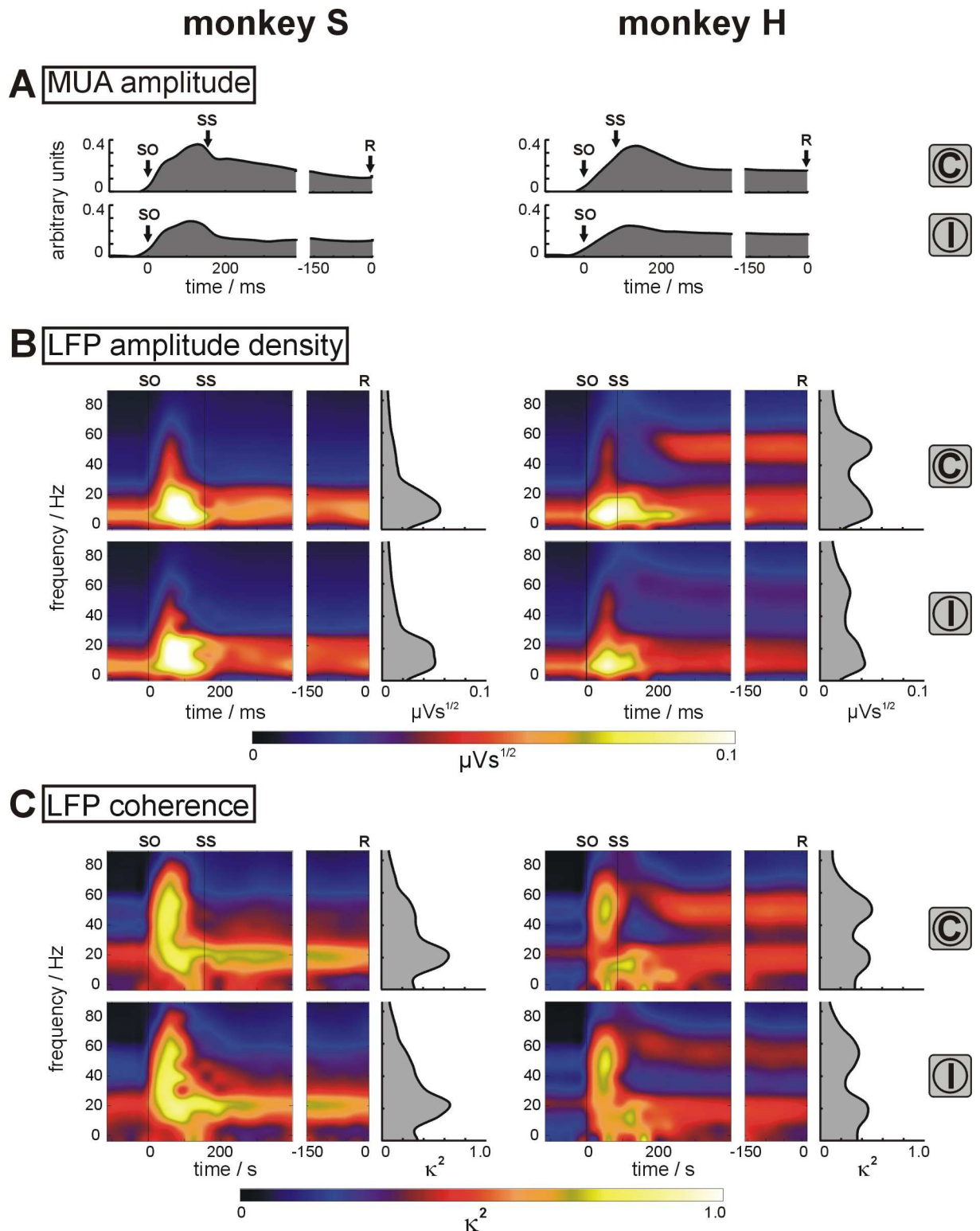
condition after decay of the stimulus-onset transient. This γ-frequency sidelobe is weaker with incongruent stimulation. Power is also slightly weaker in the low- and medium-frequency range in the incongruent compared to the congruent condition (Fig. 4.4B, right). In monkey S differences between the two conditions are small and mainly affect the low-frequency range.

Figure 4.4 General neuronal response characteristics for congruent vs. incongruent stimulation. Average MUA amplitude (**A**) LFP spectral amplitude density (**B**), and LFP pairwise spectral coherence (**C**) of all recording sites in monkey S (left 3 columns; N = 119 sites, 734 pairs) and H (right, N = 160/1086) in the congruent (Ⓒ) compared to the incongruent (Ⓘ) condition with equal left and right contrast. Within-monkey columns show data aligned with respect to stimulus-onset **SO** (left), and to behavioural response **R** (middle), respectively. Time-averages over the depicted response-triggered data is added (right) for spectral data (B, C). Differences between conditions, resulting from the stimulus switch (**SS** marks average time of switching) from piecemeal to congruent stimuli 50–250 ms after stimulus onset in the congruent condition, can be seen in the stimulus-triggered averages of all measures. Additional differences between the monkeys and the conditions can be seen in the spectral data, mainly affecting the γ-range (28–90 Hz): Monkey H shows a sidelobe between 40 and 60 Hz in power and coherence, which is not visible in monkey S. Overall in this grand-average data there is larger γ-frequency power and coherence with congruent compared to incongruent stimulation. Note that for comparability the MUA amplitude data is subjected to an equivalent sliding window procedure as the spectral data, which accounts for the smoothness.

Note that although spectral power on average peaks at low frequencies, relative enhancement compared to prestimulus baseline is broad-band, mainly across the medium and high-frequency range during the sustained epoch of the responses (>200 ms). The corresponding MUA spectra look similar but with less pronounced low-frequency power (data not shown). Pairwise inter-electrode coherence of LFP shows no apparent difference at low and medium, and a weak difference at high frequencies between the two conditions in both monkeys (Fig. 4.4C). With the given inter-electrode spacings (0.75–3.2 mm) corresponding MUA coherences are mostly too low to be considered adequately. Differences in spike density (MUA amplitude) between the two conditions are restricted to effects resulting from the slightly distinct stimulation protocol, i.e., the inserted stimulus switch in the congruent compared to the incongruent condition (Fig. 4.4A).

4.3.4 Perception-related LFP power

LFP power showed the most prominent results of all considered measures. Many recording sites showed perception-related modulations in LFP power around the time of the monkey's decision. Figure 4.5 gives an example of consistent modulations in the congruent and incongruent conditions, respectively. The modulations are consistent in the sense that higher power results in both conditions from perception of the same orientation, here horizontal. The difference in spectral amplitude does not equally affect the entire frequency range. In the congruent condition power during the percept of the horizontal grating is enlarged compared to vertical below 30 Hz, but not above. In the incongruent condition the power increase is confined to the range 10–40 Hz with a maximum around 20 Hz. The given example underlines the distinctiveness between different frequency ranges. Only the medium-frequency range (12–27 Hz) fullfills the criterion for *perception-related* modulations at this recording site (cf. Methods). The
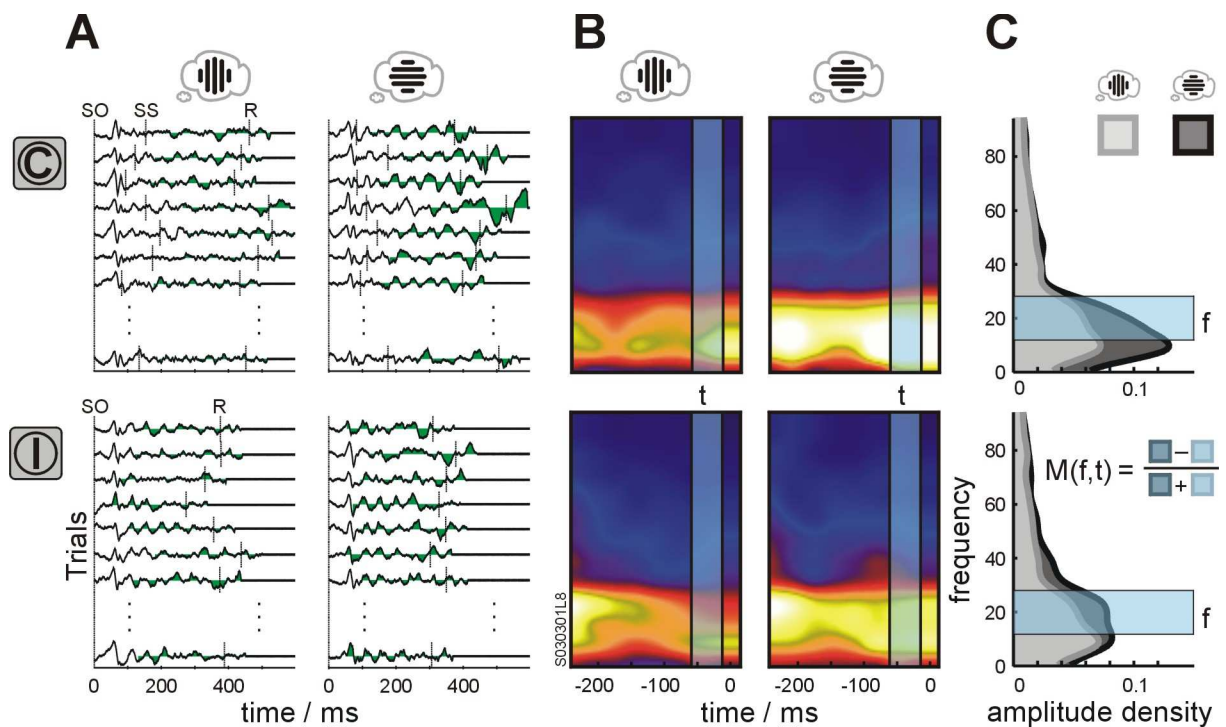
Figure 4.5 Example for perception-related modulation of LFP power. LFP data at a single recording site of monkey S is shown in 4 different conditions. Upper row: congruent (ⓒ) conditions, lower: incongruent (Ⓘ), left columns: report of vertical percept, right: horizontal. **A:** Original LFP traces (negativity up). Epochs subjected to response-triggered averages (B) are underlaid in green. **B:** Response-triggered time-frequency maps of the data in A. Transparent light blue bars indicate the time depicted as curve spectra in C. **C:** Definition of perception-related modulations. Frequencies (anywhere within the pre-defined frequency ranges), at which in both conditions there is a significant modulation with the same preference with respect to the perceived grating orientation (horizontal vs. vertical), are considered to be perception-related (cf. Methods). For the given example only the power at medium frequencies (12–28 Hz; transparent light blue bar) fullfills this criterion.

low-frequency range (4–12 Hz) is *stimulus-related*, but not *perception-related*. The high-frequency range (28–90 Hz) shows no preference in this example.

In the population data perception-related modulations are more common in the LFP low-frequency range than in other frequency ranges or in MUA. Fig. 4.6 contrasts modulation indices of MUA amplitude (6A, upper panels) and LFP low-frequency power (6A, lower) for all recording sites of monkey S in the incongruent and congruent condition for different time epochs. This comparison reveals clear distinctivness between MUA amplitude and LFP low-frequency power with respect to perception-related modulations: While modulation indices for low-frequency LFP show increasing numbers of perception-related channels towards the time of decision, MUA indices are widely scattered without discernible temporal development. At the time of decision the modulation indices in the incongruent (rivalrous) condition are broadly and symmetrically distributed around zero in MUA, while many low-frequency LFP indices are shifted towards the positive bisection line, indicating consistent modulations in both conditions, and hence perception-related modulations. Note that at the time of decision

there is no case of LFP low-frequency power significantly modulated in the incongruent condition, but in the opposite direction of the modulation in the congruent condition (*anti-modulation*). For monkey S also LFP γ-frequency power shows a tendency to perception-related modulations, but tests based on single channel data revealed no significances. Besides, in monkey H the γ-frequency power does not show this tendency for perception-related modulations, while modulations of MUA amplitude and LFP low-frequency power look similar in both monkeys (single channel data not shown).

4.3.5 Population data

For quantitative evaluation of the population data, only those recording sites are taken into account that showed orientation specific modulation in the congruent condition for the measure in consideration. Population data is quantified in two ways. First, the number of single recording sites significantly and consistently modulated in both conditions is counted (Fig. 4.6, 4.7 A). Second, because for some measures there is a tendency towards consistent modulations in the congruent and incongruent conditions (although not reaching significance in the single channels), we calculated for all recording sites significantly modulated in the congruent condition the average ratio of the modulation indices in these two conditions (*co-modulation ratio*: $\text{CMR} = 1/\text{N} \sum M_{\text{incongruent}} / M_{\text{congruent}}$). Being greater than zero this **CMR** value indicates the tendency towards perception-related modulations in the population data (Fig. 4.7B, details see below).

4.3.6 Dynamics of perception-related modulations

The number of perception-related recording sites in LFP low-frequency power has been shown to increases towards the time of decision (Fig. 4.6A, lower). At around 128 ms prior to decision the modulation indices in the incongruent condition essentially scatter around zero (although many channels are significantly modulated in the congruent condition). At the time of decision the number of perception-related channels reaches 34 in this monkey, while no anti-modulated channels are present. The corresponding data for multi-unit spike rate (MUA amplitude) does not show this temporal development: The numbers of perception-related and anti-modulated channels stay balanced (Fig. 4.6A, upper). Fig. 4.6B summarizes the temporal development of all used measures quantitatively. The time courses are similar for all measures showing perception-related modulations. The asymmetry in favour of perception-related modulations starts about 100 ms before the behavioural response of the monkeys.

4.3.7 Comparison of different measures

Fig. 4.7 summarizes the results for all measures in consideration by means of the maximal percentage of significantly modulated (consistently and oppositely) channels (Fig. 4.7A) and the maximal co-modulation ratio (Fig. 4.7B), respectively. The maximum is taken from the last 6 evaluation epochs before behavioural decision (Fig. 4.6B). Both monkeys show at large consistent results.

LFP power (Fig. 4.7A, B; columns 2–4) reveals a graded effect for the different frequency ranges, with perception-related modulations predominantly found at low frequencies, and moderately present at medium frequencies. With respect to single channel significances the high-frequency range is not significantly perception-related. In contrast, the CMR ex-
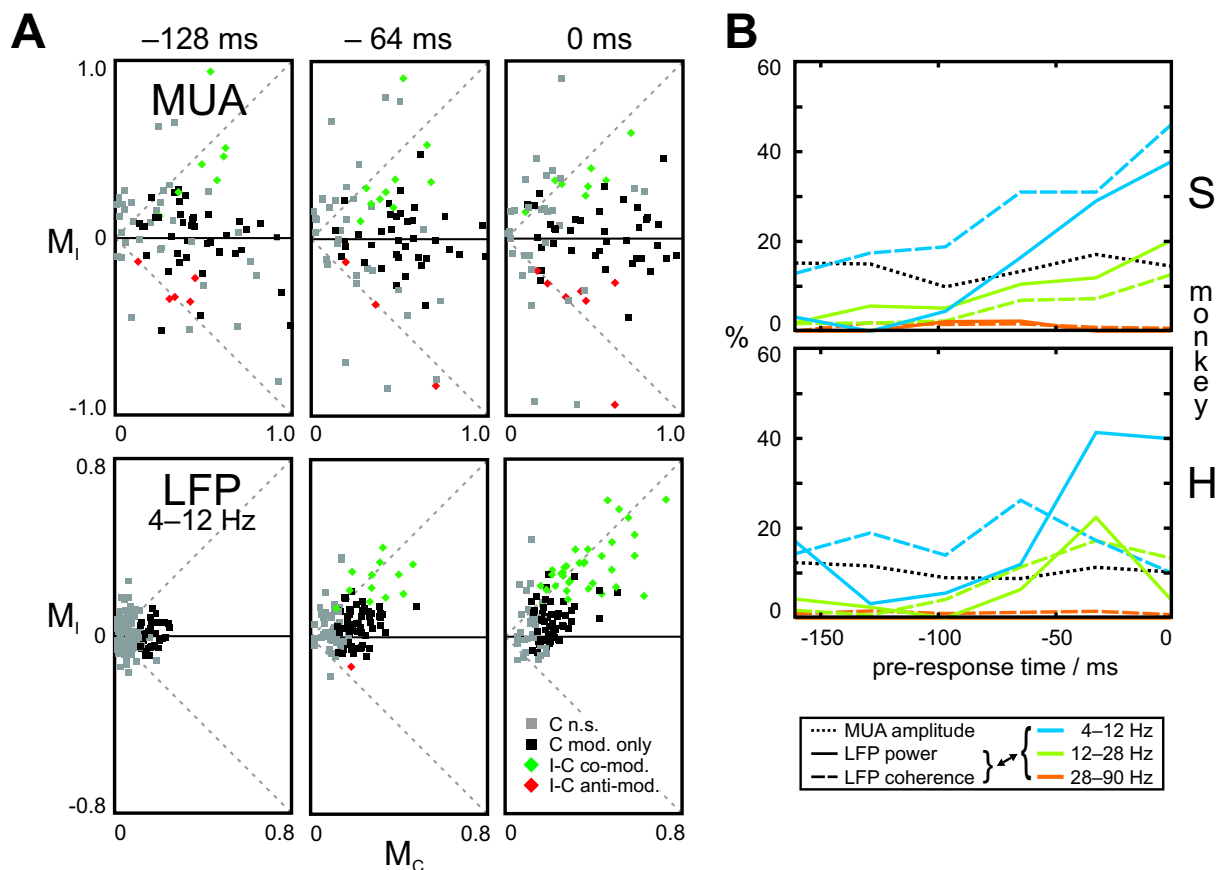
Figure 4.6 Dynamics of perception-related modulations. **A:** Examples of two different measures in monkey S. Comparison of MUA (upper row) and corresponding LFP low-frequency power (lower) results examplified by means of all recording positions of monkey S at three different times relative to the behavioural response. Modulation indices in the incongruent condition (M_I) are plotted against modulation indices in the congruent condition (M_C) for each recording position and epoch. Grey: no significant modulation in the congruent condition; black: modulation only in the congruent condition; green/red: modulation in both conditions with equal/opposite sign. Green means perception-relatedness according to our definition (cf. Methods and Fig. 4.5). **B:** Overview of temporal development of perception-related modulations for different measures. The curves show the percentage of perception-related channels compared to all channels modulated in the congruent condition for the last 6 epochs before the behavioural responses of the monkeys. Note that the incline on average is restricted to the last 100 ms before response.

Figure 4.7 Overview of population data on perception-related modulations for different measures. **A:** Maximal percentages of perception-related recording sites for the different measures in consideration (green) and corresponding numbers of anti-modulated sites (red) of the epochs prior to or at the time of response. Black numbers denotes the number of recording sites significantly modulated in the congruent condition for each measure. **B:** Maximal co-mo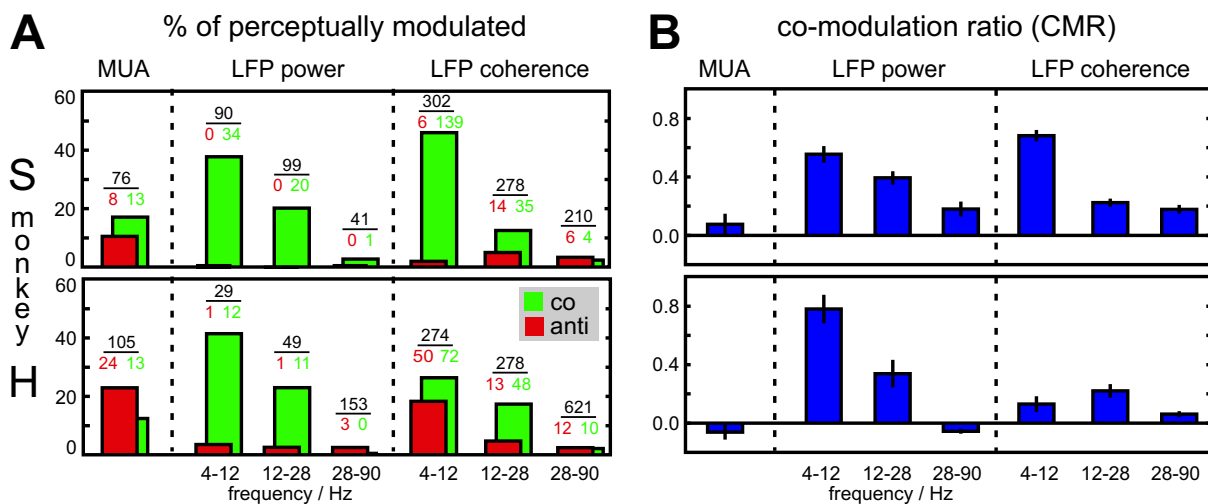dulation ratio (cf. Results) for the different measures (error bars: SD of mean). **A, B:** The results are at large consistent between both monkeys. A general trend is that perception-related modulations occur more probable at low-to-medium frequencies (4–28 Hz) and are almost absent at γ-frequencies (28–90 Hz), although significant modulations are often found at γ-frequencies with congruent stimulation (especially in monkey H).

presses a small asymmetry in favour of perception-related modulations, but only in monkey S. Note that in both monkeys there are many recording sites modulated in the high-frequency range in the congruent condition, but not in the incongruent one. Especially in monkey H, which showed a substantial sidelobe in the γ-range in the overall average (Fig. 4.4B, right), there are much more channels in the high- than in the low- or medium-frequency range solely modulated in the congruent condition, i.e., stimulus-related channels according to our definition.

LFP coherence (Fig. 4.7A, B; columns 5–7) shows in monkey S similar results as LFP power in this monkey. Especially the low-frequency coherence often is perception-related. Medium and high frequencies show moderate effects. Low-frequency coherence reveals the only major difference between the monkeys, given by the fact that in monkey H there are many (50/274) anti-modulated cases, hence pulling the CMR below 0.2. Therefore, for monkey H coherence and power differ in frequency specificity: Coherence is only in the medium-frequency range moderately perception-related.

MUA amplitude reveals in both monkeys several instances of perception-related modulations in the sense of the above given definition: $P = 17\%$ (13/76) in monkey S, $P = 12\%$ (13/105) in monkey H (Fig. 4.7A, 1st column). But in contrast to LFP power, here the numbers of anti-modulated recording sites are about the same height: S 11 % (8/76), H 23 % (24/105).

Accordingly, the mean ratio of modulations in incongruent vs. congruent condition (CMR) is nearly zero (Fig. 4.7B, 1st column).

Spectral decomposition of the MUA signal did not yield any frequency specific clustering of perception-related modulations either (data not shown). MUA spectral coherences for most electrode pairs did not exceed the to be expected bias (cf. Methods) and hence were not further analysed.

4.3.8 No difference in ocular dominance of perception-related recording sites

The ocular dominances of those recording sites revealing perception-related modulations do not differ in strength (by means of the monocularity index MI) from the whole sample (rank-sum test, $p > 0.1$) in both monkeys. There is no difference between the subset of perception-related and those solely modulated in the congruent condition (stimulus-related) either ($p > 0.1$).

One might argue, that ocularity of the perception-related channels in our study is not the critical measure to judge the relevance of inter-ocular rivalry. Perception-related modulations here are qualified by means of selectiveness for perceived orientation. A hypothetical monocular channel with arbitrary orientation selectivity, being switch on and off according to the current eye dominance or suppression (but not according to dominance/suppression of the channel's preferred orientation), would not show any modulation here, since stimulus orientation was randomly interchanged between left and right from trial to trial.

We therefore did the whole analysis in an alternative way: We used monocular stimulation as the reference condition in this case, i.e., the modulation index for the 'congruent' condition was now calculated between left and right monocular stimulation, irrespective of the grating orientation. We then sorted all incongruent trials with respect to the side of the perceived stimulus ('decision for right' vs. 'decision for left'). This leads to psychometric performance curves in complete analogy to Fig. 4.3, but this time depending on the stimulus contrast between right and left (instead of horizontal and vertical). To check for perception-related modulations depending on the side of the perceptually dominant stimulus, we again only compared trials with equal left/right contrast. This procedure did not yield a population tendency in favour of perception-related ocular selectivity in any of the used measures (data not shown).

4.4 Discussion

Our results show a complex picture of V1 involvement in binocular rivalry. This includes a major difference between multi-unit spike rate and LFP power. While LFP in many cases is modulated in consonance with alternations in the monkeys perceptual state, this is not the case for multi unit spike rate. The main result of this investigation is that significant perception-related modulations of LFP power and synchrony among pair recordings are present in V1 at low and medium frequencies, but, against expectation, are weak or absent at γ-frequencies. Perception-related modulations occured with respect to orientation but not to ocular preference.

4.4.1 Psychophysical performance

With binocular rivalry we took great care in designing the task and conducting the training to ensure that the monkeys' reports reliably reflected perception. We are convinced that both monkeys experienced binocular rivalry and correctly reported their percepts because of four reasons: 1) They reported correctly in the catch trials. 2) Their psychometric function is nearly identical to those of humans. The smooth transition in the probability of seeing either of the stimuli in dependence of the relative stimulus contrast can hardly result from any other strategy (e.g., random behaviour in the case of incongruent stimuli) than honestly reporting the percept. 3) Reaction times are slower around zero contrast difference. When flashing incongruent stimuli for less than 150 ms, they appear fused in a plaid like manner to most human observers (Wolfe, 1983). This means, exclusive dominance of one of the two simuli is retarded after stimulus onset, and therefore one should expect later reports of the stimulus orientation in the strongly rivalling compared to the non-rivalling conditions. The observed delays of our monkeys' responses fit the human data. This fact also necessitated the introduction of a piecemeal epoch at the beginning of the congruent catch trials (cf. Methods). 4) The probability of piecemeal reports peakes at zero contrast difference. This fits expectancy equivalently to the slowed reaction times and was also observed in our human data.

Taken together, these results indicate that the monkeys followed the strategy meant to. It remains unclear however, how strict the monkeys' criterion for 'exclusive dominance' was. The very low probabilities of piecemeal reports compared to our human subjects may indicate a less strict criterion, hence a disposition to indicate a stimulus orientation before the rivalry is fully resolved in favour of this orientation.

4.4.2 Inconsistent MUA and LFP results

Partly, V1 seems capable of reflecting the current perceptual state of a monkey during rivalry-induced changes of perception. In our data this capability is visible in the low-to-medium-frequency range of the LFP signal (4–28 Hz), while the local spike densities (captured by the MUA-amplitude) are less clearly modulated with respect to the perceptual alternations, even though they were recorded simultaneously by the same electrodes. The few sites with significant perception-related MUA-amplitudes are balanced by an equal number of anti-modulated sites. In addition, there is no apparent temporal development in the MUA spike rate modulation at these sites in advance of the behavioural decision, as is the case for LFP power (Fig. 4.6). Therefore, although significant on the basis of single channel analysis, the MUA data could simply reflect incidental significances due to an 'arbitrary' threshold criterion inherent to statistical testing. In contrast, LFP modulations show a marked asymmetry in favour of perception-related modulations at the same time.

The differential results between LFP and MUA could result from a signal-to-noise advantage in LFP, if the perception-related modulations are carried by the synchronized activity of large cell populations. Similar quantitative differences between MUA and LFP results appeared in earlier studies using the same techniques (Gail et al., 2000). LFP and MUA signals are different with respect to their electrophysiological origin (LFP reflecting somato-dendritic potentials, compared to spike output in MUA), and with respect to their extracellular range of integration, i.e., the number of neurons directly contributing (Legatt et al., 1980; Mitzdorf, 1987). This has ambivalent consequences for the specifity of LFP: The somato-dendritic potentials should partly represent the orientation-selectivity of the local lateral connectivity (Gilbert, 1993), which leads to a signal-to-noise advantage in LFP compared to the more stochastic spike outputs. On the other hand, the larger range of unspecific extracellular integration over neurons of different orientation preferences contributing to LFP should reduce the degree of its orientation-selective modulations, since orientation selectivity becomes weaker with unspecific averaging, which would be counterproductive for the observed perception-related modulations. In conclusion, we consider the contribution of sub-threshold potentials (specifically integrating neurons with similar orientation preferences) to LFP responsible for the difference to the MUA results in this study.

4.4.3 Perception-related synchronization

Increased synchronization in the γ-range (35–90 Hz) among local cell populations, representing features of one of the rivalling objects, has been proposed to represent perceptual dominance of this object during rivalry (Kottman et al., 1996). Observation of increased γ-synchrony associated with perceptual dominance in a rivalry task with strabismic cats has been taken as support for this idea (Fries et al., 1997; Fries et al., 2002), although transferability to mechanisms of rivalry in normally raised animals with an intact visual system is questionable. Modulation of power in MEG recordings in correlation with the percept during rivalry has also been interpreted in favour of the synchronization hypothesis, since power in the MEG signal is associated with the degree of synchronization in local cell populations (Tononi et al., 1998). In analogy, this has been done for synchronization of large cell populations that are distributed in different cortical areas, based on perception-related large-scale MEG coherence modulations during rivalry (Srinivasan et al., 1999). Due to the method of frequency-tagging, these MEG studies, however, can give no answer on the frequency-specifity of the observed modulations, since only the power at the tag frequency is analysed.

In the present study we did not reliably find significant γ-frequency specific modulations related to perceptual state. Neither power at single electrodes nor inter-electrode coherence did show major dependence on perception in the γ-band. This seems to be in contradiction to another study from our laboratory, in which synchronized γ-activity occurred with the correct perception in a difficult figure-ground task in V2 (Wölbern et al., 2002). Note, however, that in the present study stimulus-dependent modulations of γ-frequency power and coherence did occur in the congruent stimulation condition, like they did in many previous studies (e.g., Frien and Eckhorn, 2000; Frien et al., 2000; reviews: Eckhorn, 1999; Gray, 1999; Singer, 1999). This means that in our present investigation γ-frequency power and coherence in V1 better reflect the stimulus properties than the perceptual state. In this context it is noteworthy that there is a reduction of LFP power and coherence, especially during the late response, with incongruent compared to congruent stimulation (obvious above 30 Hz; Fig. 4.4).

Nevertheless, there are perception-related modulations of LFP coherence in the low- and medium-frequency ranges. Coherence at low frequencies (theta/alpha: 4–12 Hz) has recently been discussed in terms of long-range integrative processes (e.g., von Stein and Sarntheim, 2000), based on findings that this frequency range is capable of mediating context dependent top-down feedback to primary visual cortex in a cat go-/no-go task (von Stein et al., 2000). The authors suggest couplings among cortical areas in the alpha-to-theta range to be

most pregnant, when the brain is "generating [a] hypothesis about the environment" (von Stein and Sarntheim, 2000, p. 311), especially, when bottom-up information is insufficient for a unique percept. This is also the case in our BR experiments in which sensory information is present, but ambiguous, so that the decision in favour of one or the other interpretation of the stimulus can solely be made internally, i.e., the perception-related modulations in the 4–12 Hz band may be interpreted as a signature of internal hypothesis generation about the stimulus. Another necessity for long-range integration associated with low-frequency signals in our task may results from the fact that the receptive fields of our recording sites were all close to the vertical meridan. The stimulus patches in consequence covered right- and left-side visual field positions simultaneously. The cortical representation of the whole stimulus therefore requires integration of cortical activity from both hemispheres. On the basis of our data we cannot decide, however, whether the local low-frequency activity in our data is associated with corresponding extrastriatal or contra-hemispherical activity.

4.4.4 Binocular rivalry in V1

How can the role of V1 in binocular rivalry be seen on the background of the available findings? Earlier reports on rivalry-induced modulations along the visual cortical pathway were disparate about the role of V1. Spike rates in macaque V1 of very few cells did show perception-related modulations, while with increasing level of the visual areas higher numbers of modulated cells were found (Leopold and Logothetis, 1996; Sheinberg and Logothetis, 1997). Corresponding fMRI BOLD measures in humans did show V1 activation being correlated with perception during rivalry, but not as strong as during the non-rivalrous reference condition (Polonsky et al., 2000). BOLD differences due to alternating perceptual states in humans, being as large during rivalry as during congruent viewing, were found in the blind spot representation of V1 (Tong and Engel, 2001). All these results, together with ours, would fit into the same overall picture of V1 participating in BR, given that the BOLD signal represents mostly (subthreshold) somato-dendritic activity. A recent study confirms this view (Logothetis et al., 2001). Although indirect evidence suggests that the BOLD signal well correlates with overall spike activity (V1: Heeger et al., 2000; V5: Rees et al., 2000), a direct comparison of simultaneously recorded fMRI- and microelectrode-signals showed higher correlation of BOLD with LFP than with spike density (Logothetis et al., 2001). This could explain the various results on V1 modulation during binocular rivalry, and would lead to the suggestion that these modulations are mainly subthreshold.

4.4.5 Latency of modulations

The perception-related modulations we found, appeared on average shortly before the monkeys' decisions. The first asymmetries in favour of consistent modulations in the population data can approximately be seen in the epoch centered around 96 ms (–156 to –32 ms) before the time of decision, with increasing strength towards the time of decision (only modulation of low-frequency coherence in monkey S is occurring earlier, Fig. 4.6). From the mean reaction times in the congruent conditions (monkey H: 395 ms; S: 372 ms) one can estimate the maximum interval from stimulus specific V1 activation to the motor output to about 270–320 ms (by subtracting 70–100 ms stimulus–response latency to supragranular layers in V1 (Fig. 4.5; cf. also Nowak and Bullier, 1997) from the reaction times). This estimate is conservative in two ways. First, most trials are shorter than the mean reaction time, since reaction time is typically asymmetrically distributed with a long-tailed end towards long reaction time. Therefore the mean overestimates reaction times. Second, even when assuming that V1 activity reflects perceptual states responsible for triggering decisions, it is not clear whether the earliest activation can be made responsible for this. Nevertheless, the time remaining after the first perception-related modulations to the motor output is rather short to contain the whole preparation and performance of the motor response.

Two reasons may account for the lateness of the given effects: First, from the psychophysical data the suspicion arose, that the highly trained monkeys could have been premature in making the decision, i.e., judged the oncoming percept before perceptual dominance of one grating was exclusive. Measures reflecting the perceptual state of dominance then would be late. Since we were aiming for signal components potentially causative for the monkeys' decisions, our stimulation protocol does not allow to evaluate later time epochs (after the monkey pushed the key). Hence, we can not evaluate, whether the increase in the perceptually related modulation outreaches the time of decision. Second, perception-related modulations described in this study may not originate in V1, but instead reflect top-down influence from higher visual areas. Such feedback projections are not necessarily slower than internal V1 interactions (Hupé et al., 2001), and therefore should not automatically be attributed to higher latencies. But the possibility of feedback being responsible for the observed modulations would per se take the above temporal considerations out of effect, since the V1-to-motor-output delay is not the relevant time factor any more (more about top-down cf. below).

4.4.6 Eye versus object rivalry

Ocularity of the neurons revealing perception-related activity has attained a key role in determining the relevance of the inter-ocular competition concept of rivalry on the one hand (*eye rivalry*: Lehky, 1988; Blake, 1989; Lee and Blake, 1999; Tong and Engel, 2001), and the concept of competition among high level percepts on the other hand (*object rivalry*: Kovács et al., 1996; Logothetis et al., 1996; Ngo et al., 2000). On the background of the huge variety of seemingly contradictory psychophysical and neurophysiological findings on this issue, insight gains ground that rivalry may be disentangled at several levels of processing, depending on stimulus properties (Bonneh et al., 2001; Blake and Logothetis, 2002). Highly coherent stimuli tend to evoke object rather than eye rivalry (Bonneh et al., 2001). With respect to stimulus size and uniformity we used stimuli with high perceptual coherence, according to the definition of Bonneh et al. (2001), hence with high probability to evoke object rivalry. This is consistent with the fact that in our study there is no difference in the strength of ocular dominance between those channels showing orientation-selective modulation with perception (comparing perception of a horizontal vs. vertical grating) and the remaining recording sites; in addition, we did not find perception-related modulations with respect to ocular selectivity (comparing perception of the left vs. right stimulus). On the other hand, we used high contrasts which are more probably subjected to eye rivalry, according to the findings of Lee and Blake (1999). However, our data gives no indication on eye rivalry under the given conditions.

Interestingly, a previous multi-stage model on binocular rivalry predicts decreased synchronization of binocular neurons at V1 level during incongruent compared to congruent stimulation (Lumer, 1998). This model is not based on the eye rivalry concept and creates competitive firing patterns in higher visual areas without corresponding modulation of monocular V1 neuron activity. The competitive firing patterns are interpreted by the author as kind of perception-related modulations. Basic ideas of this model may be supported by our finding of decreased coherence during incongruent stimulation, and by reduced cross-correlation observed by others (D. Leopold, personal communication), and lacking evidence for a special role of monocular recording sites.

4.4.7 Extrastriatal impact on V1?

Where could the perception-related LFP modulations lacking corresponding spike rate effects in V1 come from? There are 3 potential sources: i) subcortical, afferent input from LGNd, ii)

intrinsic, lateral V1 connections, and iii) extrinsic feedback projections from higher visual areas.

The first alternative seems to be ruled out, since LGNd in awake monkey does not show differential activation due to congruent and incongruent stimulation, respectively, and therefore is assumed not to take part in binocular rivalry (Lehky and Maunsell, 1996). Unfortunately, in these experiments the monkeys' perception was not monitored. Besides this, afferent inputs from LGN to V1 are typically excitatorily driving and not mainly modulatory.

If the second alternative was in effect, we would have to expect qualitatively consistent results in our MUA and LFP data. We typically recorded from the upper V1 layers 2/3, where intrinsic lateral feedback mainly originates from within the same layers via horizontal collaterals of pyramidal neurons (Rockland and Lund, 1983; Yoshioka et al., 1996). But then our MUA signals contain not only spike activity of those neurons being modulated, but also of those being the source of the modulatory activity. Therefore, if perception-related LFP modulations were mediated by lateral feedback impact, this information should also be contained in V1 layer 2/3 spike activity. Moreover, if binocular rivalry is in any way resolved at the level of V1 and this information is transmitted to higher visual areas, then this should be measurable in spike activities of layer 2/3, since these are the main feed-forward output layers to higher cortical levels (e.g., Rockland and Pandya, 1979). However, we could not demonstrate this by means of multi-unit spike rate, nor by MUA power in any frequency range. Due to the large stochastic components MUA coherence was in general to weak to be evaluated adequately. Of course, we can not exclude, that the information is present in some other measure of layer 2/3 spike activities. Coupling measures in the temporal domain (like cross-correlation), or multi-channel methods taking into account all recorded channels simultaneously in order to improve sensitivity (Kremper et al., 2002), could yield additional aspects of the role of V1 spike activity in binocular rivalry.

Taken together, this indirectly suggests the third possibility, feedback of perception-related signals from extrastriate cortical areas. This seems plausible, since many extrastriate feedback connections terminate in layer 2/3 of V1 (in addition to infragranular layers) and are typically modulatory in nature (reviewed in: Felleman and van Essen, 1991; Salin and Bullier, 1995). Functional imaging studies on binocular rivalry reported several extrastriate areas modulated by perceptual alternations. These include the fusiform face area (FFA) and the parahippocampal place area (PPA), when using rivalry between faces and houses (Tong et al., 1998). Early ventral stream (V1–V4) was shown to reflect perceptual alternations with the less

complex grating stimuli (Polonsky et al., 2000). Further, occipito-temporal involvement in expression of different perceptual states during rivalry was reported by Lumer et al. (1998). In this study, specific activation in fronto-parietal regions was associated with purely perceptual transitions (exclusively occurring without stimulus switches). Other whole-head studies with VEPs (e.g., Brown and Norcia, 1997) or MEG (Tononi et al., 1998; Srinivasan et al., 1999) lacked spatial resolution to resolve the participating cortical sites in terms of area definitions. In monkey binocular rivalry Sheinberg and Logothetis (1997) have shown that BR is fully 'resolved' in inferotemporal cortex, visible in single cell spike rate modulations being almost always in concordance to perceptual switches. In summary, influence by feedback from temporal areas on V1 is, from the known direct (Rockland and van Hoesen, 1994) or indirect (Felleman et al., 1997) projections, conceivable.

4.4.8 Summary and Conclusion

The activity in striate cortex (V1) of monkeys is modulated in consonance with the alternative percepts in binocular rivalry. Perception-related modulations occurred in power and coherence of LFP at low-to-medium frequencies, but not in the γ-range of LFP, or any frequency range of MUA. This demonstrates that even in V1 activity components appear that are correlated with visual awareness. The question remains, in how far these components represent awareness-relevant activity itself, or take part in an awareness-relevant selection process, or just are an epiphenomenon of other awareness-relevant processes, not captured in the present recordings. The somato-dendritic nature of the perception-related components (LFP), together with their close temporal appearence to the behavioural response, make it questionable for being causative for the perceptual switch itself. Instead we argue for a feedback impact on V1 from other cortical areas already representing or carrying a prediction about the oncoming perceptual state. The role of such an impact could be to stabilize a newly established percept by supporting V1 neurons in representing features of the currently or oncoming dominant stimulus. It would be tempting to know, whether this can lead to fast synchronization of γ-activity in V1 according to the feature-binding hypothesis. To test such an idea, it would be helpful to make the same investigation with recordings lasting over the time of several perceptual switches, to be able to observe signals during longer phases of stable percepts.

5   Concluding Remarks

The main experiments of this study addressed two important aspects of object perception, figure–ground segregation and visual awareness. In both experiments the role of synchronisation for feature binding across space in the retinotopically organized primary visual cortex was matter of subject. Detailed discussion of the results can be found in the respective chapters, but a few summarizing remarks may be helpful[1].

Synchrony and visual segmentation. The outcome of the *figure–ground experiment* (Chapter 2) was consistent with our expectation; decoupling across segment borders is predicted by the binding-by-synchronisation hypothesis.

Accompanying network simulations show how this decoupling can be achieved with physiologically plausible assumptions on the connectivity and lateral interaction within one feature dimension (orientation of local contrast-borders) organized in a retinotopical map as typically found in V1 (Saam et al., 2000; Al-Shaikhli, 2001). Discontinuity of the relevant feature (not necessarily feature contrast) within a topographic representation turned out to be a sufficient prerequisite for decoupling across segments in the models.

Across object contours in real world often multiple features reveal a discontinuity (e.g., different stereoscopic depth and different colour of figure versus background). But a feature may also continue unchanged across object borders; the principle of camouflage tries to minimise discontinuities (e.g., the striped texture on the fur of a tiger in the prairie). The distinction on whether or not parts of a scene become segregated may reflect the net balance of continuities and discontinuities across all basic feature dimensions at a potential segment border. (If the tiger stands still we might oversee him because of the continuity of colour and texture; if he moves there is an additional discontinuity in the visual field with respect to motion, which helps to segregate him from the background.) Within each feature dimension spatial neuronal (de-)synchronisation of fast signal components may adjust on a retinotopical map according to the (dis-)continuity of the feature in space. The net (integrating over all elementary feature dimensions) state of synchronisation along space could then decide about perceptual segregation. Primary visual cortex V1 is predestined to house such a mechanism since multi-

---

1   especially for those who skipped the previous chapters ☺.

ple features are represented within the same retinotopy. It is tempting to test this segmentation hypothesis with a more comprehensive class of figure–ground stimuli than used in the experiments presented here.

Synchrony and visual awareness. The results of the *binocular rivalry experiment* (Chapter 4) were not consistent with our expectation. What does this mean for the binding-by-synchronisation hypothesis?

The hypothesis states synchronisation for the purpose of binding features into objects. It was proposed that this binding is a necessary prerequisite for perceiving an object, and may be closely related to visual sensory awareness (Crick and Koch, 1990; reviews: Crick and Koch, 1998; Engel and Singer, 2001). But is binding sufficient for *conscious* perception? Can objects subconsciously become bound to form coherent entities without entering a state of visual awareness? One of the current views on the perceptual phenomenology of binocular rivalry suggests "yes"[2]. The *object rivalry* approach assumes coherent objects at high levels of processing to compete for access to awareness in binocular rivalry (cf. Paragraph 4.4.6). This means, there would have to be subconscious neuronal representations of both entire objects. If these representations, including their neuronal prerequisites at early processing stages like V1, are untouched by perceptual switches, then V1 would be ruled out for contributing causatively to an awareness-relevant selection process.

In the binocular rivalry experiment the neuronal synchronisation was tested for correlation with the behavioural reports of the monkeys, which are assumed to reflect their momentary state of conscious perception. The results with respect to γ-synchronisation could still be consistent with the binding hypothesis, if the binding processes carried by temporal signal correlation are seen as *pre-conscious analysis* of the informational content of the sensory stimulus which regard rules of perceptual organization (e.g., with respect to segmentation), but not necessarily lead to visual awareness. To test such a hypothesis, it will be necessary to investigate the states of γ-synchronisation found with rivalrous (left–right incongruent) stimulation in more detail. The results with respect to synchronisation at low-to-medium frequencies eventually may be interpreted in the sense of top-down interaction (hierarchical binding) coming into play in ambiguous viewing situation, where the sensory stimulus in not sufficient to induce a

---

2   There has been plenty of other research on the relation between subconscious and conscious vision, and the putative role of different visual subsystems contributing to it, based, for example, on lesion studies, and including phenomena like blindsight (e.g., Stoerig and Cowey, 1997; Stoerig et al., 2002); some recent reviews on visual awareness are: Stoerig, 1996, 2001; Crick and Koch, 1998; Pollen, 1999; Zeki and Bartels, 1999; Lamme et al., 2000, Engel and Singer, 2001.

unique perceptual state, and embedding of the sensory information into higher-order contexts makes additional interplay between several processing stages necessary.

Phase-continuity for object-continuity? Neuronal synchronization at γ-frequencies in the cortex, as considered here, is on average limited in its spatial extent to a few millimetres, even within regions representing coherent object surfaces across a wider cortical range (Jürgens et al., 1996; Frien and Eckhorn, 2000). The observed scopes of synchronization allow to expand the classical receptive field concept to the notion of association fields (reviews: Eckhorn, 1999, 2001), but make synchronization, strictly speaking, a questionable candidate for coding object coherence over its entire range.

However, more sophisticated analyses of the spatial distribution of signal coherence in the identical recordings suggested previous findings at least partly to be a technical consequence of averaging: The temporal variability of relative phase between two signals has been shown to increase proportionally with the distance of the two recording sites. But for a single point in time the phase gradient along space (determined with multiple aligned recording sites) is often found to be constant. Based on these observations it is proposed that object-continuity may be coded by *phase-continuity* (Gabriel and Eckhorn, 1999; Eckhorn et al., 2001) with variable phase gradients, rather than by strict synchrony. For target neurons at a subsequent processing stage, each with a confined region of support from the previous stage, phase-continuity coding would effectively not make much difference to strict synchrony, as long as the phase gradients are sufficiently small that temporal delays across the region of support for the target neurons do not exceed their finite temporal range of integration. Indeed, model considerations suggest that, assuming Hebbian learning, the 2nd stage receptive field sizes are determined by the width of the 1st stage association fields (Saam and Eckhorn, 2000), i.e., the regions of support for the 2nd stage RFs do not exceed the scope of synchronization at the first stage, which is smaller than the scope of phase-continuity. The results of the figure–ground experiment are compatible with this broadened view on neuronal segmentation for object coding.

Need for speed? The binding-by-synchronization hypothesis has been challenged in recent reports relying on the psychophysical finding that neuronal synchrony *externally* forced by flickering visual objects is not supportive for scene segmentation (e.g., Kiper et al., 1996; Fahle and Koch, 1995; but see, e.g., Leonards et al., 1996; Blake and Yang, 1997).

Fast transient stimulus changes, as in flickering stimulation, evoke early neuronal responses phase-locked to the onset of stimulus change (~30–80 ms latency in V1), while later activities during sustained visual stimulation and ocular fixation are likely to reveal intrinsic synchronization, often of oscillatory character, and not phase-locked to the onset of stimulus change (e.g., Eckhorn and Schanze, 1991; Jürgens et al., 1999). Phase-locked response transients due to fast stimulus changes or (micro-)saccadic eye movements disturb stimulus-induced intrinsic fast oscillations (Eckhorn and Schanze, 1991; Kruse and Eckhorn, 1996; Güttler et al., 1996). In the figure–ground experiment, presented here, the stimulus specificity of synchronization (i.e., the decoupling across the object contour) was visible only in the late stimulus-induced response epochs, but not in the early stimulus-locked components.

Does this, together with the above mentioned psychophyhsical results, suggest a notion of only the late (> 100 ms latency) response components, not phase-locked to external stimulus events, being capable to represent scene segmentation? Such a notion would have to be reconsidered to go well with the very fast categorization capabilities in human and monkey vision interpreted in favour of fast feed-forward processing (e.g., Thorpe et al., 1996; Delorme et al. 2000). Besides, observations from different studies revealed a broad capability of the visual system to manage scene segmentation based on temporal cues, ranging from evaluation of short temporal stimulus onset asynchronies (5–10 ms; e.g., Fahle, 1993; Leonards et al., 1996) to long-lasting temporal integration for figure perception (several ten to hundred milliseconds; e.g., Altman et al., 1986; Fahle and Koch, 1995). The relation of fast synchronization to such different demands on velocity in visual processing is an unsolved issue.

Answering open questions, like the consequences of the limited spatial range of cortical synchronization for coding large objects, the temporal characteristics of synchronization effects on the background of very fast perceptual segmentation and categorization capabilities, or their relevance for visual awareness will be essential to pin down the role of synchronization in object perception in future research.

Finally, I again want to emphasize that different concepts for neuronal processing, like feed-forward filtering, the mechanisms attributed to contextual modulations or selective attention, or the synchronisation for feature binding, should not be understood as mutually exclusive in the visual system. The idea of feature binding by synchronisation, for example, is based on the assumptions that assemblies constitute of participating neurons with intermediate-level filter properties. And it does not exclude the possibility that the informational content carried by an

assembly is selectively and flexibly projected onto highly specific neurons or neuronal assemblies. This view would assign synchronisation a place at some intermediate level of processing to allow flexibility when necessary, but complementary to established hard-wired processing for meeting permanent requirements in a fast and robust manner. The flexibility of synchronisation, on the other hand, could well be used by contextual or attentional mechanisms to guide selection processes. But these are speculations our brain gives plenty of room for – both, as a subject of research and as the executing organ.

Epilogue

To pick up the preface's introducing pun one may ask who the *blind man* is. In view of the present still poor state of knowledge in visual neuroscience, maybe we as scientists should feel addressed. But there has also been a lot of promising progress in understanding the basics of vision. And after all it holds that *In the kingdom of the blind the one-eyed is king*. So, it all depends on how you look at it. Let me therefore close with an sceptical yet optimistic *We are the blind men, but we will see*.

References

Alais D, O'Shea RP, Mesana-Alais C, Wilson IG (2000) On binocular alternation. Perception 29:1437-1445.

Al-Shaikhly B (2001) Figur/Hintergrund-Trennung durch Signalentkopplung. Diploma Thesis. Physics Department. Marburg University.

Altmann L, Eckhorn R, Singer W (1986) Temporal integration in the visual system: Influence of temporal dispersion on figure-ground discrimination. Vision Res 26:1949-1957.

Andrews TJ (2001) Binocular rivalry and visual awareness. Trends Cogn Sci 5:407-409.

Anzai A, Ohzawa I, Freeman RD (1999a) Neural mechanisms of encoding binocular disparity: Receptive field position versus phase. J Neurophysiol 82:874-890.

Anzai A, Ohzawa I, Freeman RD (1999b) Neural mechanisms for processing binocular information I. Simple cells. J Neurophysiol 82:801-908.

Anzai A, Ohzawa I, Freeman RD (1999c) Neural mechanisms for processing binocular information II. Complex cells. J Neurophysiol 82:909-924.

Ashby FG, Prinzmetal W, Ivry R, Maddox WT (1996) A formal theory of feature binding in object perception. Psychol Rev 103:165-192.

Barlow HB (1972) Single units and sensation: A neuron doctrine for perceptual psychology? Perception 1:371–394.

Barlow HB, Blakemore C, Pettigrew JD (1967) The neuronal mechanisms of binocular depth discrimination. J Physiol 193:327-342.

Benignus VA. [title unknown]. IEEE Trans Aud and ElectroAcoust 17:145-150.

Blake R (1989) A neural theory of binocular rivalry. Psychol Rev 96:145-167.

Blake R, Logothetis NK (2002) Visual competition. Nat Rev Neurosci 3:1-11.

Blake R, Yang Y (1997) Spatial and temporal coherence in perceptual binding. Proc Nat Acad Sci USA 94:7115-7119.

Bonneh Y, Sagi D, Karni A (2001) A transition between eye and object rivalry determined by stimulus coherence. Vision Res 41:981-989.

Boucart M (1999) The neuroscience of perceptual integration. Hove: Psychology Press.

Brinksmeyer HJ, Gail A, Eckhorn R, Thomas U (2001) Dynamic correction of receptive field positions for residual eye movements in monkey V1? Soc Neurosci Abstracts 27:619.37.

Brosch M, Bauer R, Eckhorn R (1997) Stimulus-dependent modulations of correlated high-frequency oscillations in cat visual cortex. Cerebral Cortex 7:70-76.

Brown RJ, Norcia AM (1997) A method for investigating binocular rivalry in real-time with the steady-stade VEP. Vision Res 37:2401-2408.

Bruns A, Eckhorn R (2002) Task-related inter-areal coupling between high- and low-frequency signals in human visual cortex. Submitted to Int J Psychophysiol.

Chelazzi L, Duncan J, Miller EK, Desimone R (1998) Responses of neurons in inferior temporal cortex during memory-guided visual search. J Neurophysiol 80:2918-2940.

Crick F, Koch C (1990) Towards a neurobiological theory of consciousness. Semin Neurosci 2:263-275.

Crick F, Koch C (1998) Consciousness and neuroscience. Cereb Cortex 8:97-107.

Cumming BG, Parker AJ (1997) Responses of primary visual cortical neurons to binocular disparity without depth perception. Nature 389:280-283.

Cumming BG, Parker AJ (1999) Binocular neurons in V1 of awake monkeys are selective for absolute, not relative disparity. J Neurosci 19: 5602-5618.

Delorme A, Richard G, Fabre-Thorpe M (2000) Ultra-rapid categorisation of natural scences does not rely on colour cues: a study in monkeys and humans. Vision Res 40:2187-2200.

Desimone R (1998) Visual attention mediated by biased competition in extrastriate visual cortex. Philos Trans R Soc Lond B Biol Sci 353:1245-1255.

Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. Annu Rev Neurosci 18:193-222.

DeValois RL, Albrecht DG, Thorell LG (1982) Spatial frequency selectivity of cells in macaque visual cortex. Vision Res, 22:545-559.

Diaz-Caneja E (1928) Sur l'alternance binoculaire. Ann Ocul (Paris):721-731. [Translated in Alais et al., 2000.]

Dow BM, Gouras P (1973) Color and spatial specificity of single units in rhesus monkey foveal striate cortex. J Neurophysiol, 36:79-100.

Düchting H (1992) Wassily Kandinsky 1866-1944. Revolution der Malerei. Köln: Taschen.

Eckhorn R (1999) Neural mechanisms of visual feature binding investigated with microelectrodes and models. Visual Cogn 6: 231-265.

Eckhorn R (2001) Neural principles of preattentive scene segmentation: Hints from cortical recordings, related models, and perception. In: Models of Neural Networks IV (Van Hemmen JL, Cowan JD, Domay E, eds). Berlin, Heidelberg, New York: Springer.

Eckhorn R, Bauer R, Jordan W, Brosch M, Kruse W, Munk M, Reitböck HJ (1988) Coherent oscillations: A mechanism of feature linking in the visual cortex? Biol Cybern 60:121-130.

Eckhorn R, Bruns A, Saam M, Gail A, Gabriel A, and Brinskmeyer HJ (2001) Flexible cortical gamma-band correlations suggest neural principles of visual processing. Visual Cogn 8:519–530.

Eckhorn R, Frien A, Bauer R, Wölbern T, Kehr H (1993a) High frequency (60-90 hz) oscillations in primary visual cortex of awake monkey. Neuroreport 4:243–246.

Eckhorn R, Krause F, Nelson JI (1993b) The RF-cinematogram. A cross-correlation technique for mapping several visual receptive fields at once. Biol Cybern 69:37-55.

Eckhorn R, Reitböck HJ, Arndt M, Dicke P (1990) Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex. Neural Comput 2:293–307

Eckhorn R, Schanze T (1991) Possible neural mechanisms of feature linking in the visual system: stimulus-locked and stimulus-induced synchronizations. In: Self-Organization, Emerging Properties, and Learning (Babloyantz A, ed), pp63-80. New York: Plenum Press.

Eckhorn R, Thomas U (1993) A new method for the insertion of multiple microprobes into neural and muscular tissue, including fiber electrodes, fine wires, needles and microsensors. J Neurosci Methods 49:175-179.

Engel AK, Fries P, König P, Brecht M, Singer W (1999) Temporal binding, binocular rivalry, and consciousness. Consci Cogn 8:128-151.

Engel AK, König P, Singer W (1991a) Direct physiological evidence for scene segmentation by temporal coding. Proc Natl Acad Sci USA 88:9136-9140.

Engel AK, Kreiter AK, König P, Singer W (1991b) Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. Proc Natl Acad Sci USA 88:6048–6052.

Engel AK, Singer W (2001) Temporal binding and neural correlates of sensory awareness. Trends Cogn Sci 5:16-25.

Erb M, Aertsen A (1993) Dynamics of activity in biology-oriented neural network models: Stability at low firing rates. In: Brain Theory - Spatio-temporal aspects of brain function (Aertsen A, ed), pp 201-223. Amsterdam: Elsevier.

Ernst B (1992) Der Zauberspiegel des M.C. Escher. Köln: Taschen.

Fahle M, Koch C (1995) Spatial displacement, but not temporal asynchrony, destroys figural binding. Vision Res 35:491-494.

Fahle M (1993) Figure-ground discrimination from temporal information. Proc R Soc Lond B 254:199-203.

Feist PH (2000) Malerei des Impressionismus 1860-1920. Teil I Der Impressionismus in Frankreich (Walther IF, ed). Köln: Taschen.

Felleman DJ, van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1:1-47.

Felleman DJ, Xiao Y, McClendon E (1997) Modular organization of occipito-temporal pathways: Cortical connections between visual area 4 and visual area 2 and posterior inferotemporal ventral area in macaque monkeys. J Neurosci 17:3185-3200.

Ferster D (1981) A comparison of binocular depth mechanisms in areas 17 and 18 of the cat visual cortex. J Physiol 311: 623-655.

Field DJ, Hayes A, Hess RF (1993) Contour integration by the human visual system: Evidence for a local „association field". Vision Res 33:173-193.

Freeman RD, Ohzawa I (1990) On the neurophysiological organization of binocular vision. Vision Res 30:1661-1676.

Frien A, Eckhorn R (2000) Functional coupling shows stronger stimulus dependency for fast oscillations than for low-frequency components in striate cortex of awake monkey. Eur J Neurosci, 12:1466-1478.

Frien A, Eckhorn R, Bauer R, Wölbern T, Gabriel A (2000) Fast oscillations display sharper orientation tuning than slower components of the same recordings in striate cortex of the awake monkey. Eur J Neurosci, 12:1453-1465.

Frien A, Eckhorn R, Bauer R, Wölbern T, Kehr H (1994) Stimulus-specific fast oscillations at zero phase between visual areas V1 and V2 of awake monkey. NeuroReport 5:2273-2277.

Fries P, Roelfsema PR, Engel AK, König P, Singer W (1997) Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry. Proc Natl Acad Sci USA 94:12699-12704.

Fries P, Schröder JH, Roelfsema PR, Singer W, Engel AK (2002) Oscillatory neuronal synchronization in primary visual cortex as a correlate of stimulus selection. J Neurosci 22(9):3739-3754.

Fukushima K (1980) Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 36:193-202.

Gabriel A, Eckhorn R (1999) Phase continuity of fast oscillations may support the representation of object continuity in striate cortex of awake monkey - correlation analysis of time- and space-resolved single responses. In: Proceedings of the 27th Göttingen Neurobiology Conference (Elsner N Eysel U, eds), 489. Stuttgart, New York: Georg Thieme.

Gail A, Brinksmeyer HJ, Eckhorn R (1999a) Different possible contributions of striate cortex activity to visual object representation in awake monkey. In: Proceedings of the 27th Göttingen Neurobiology Conference II (Elsner N, Eysel U, eds), p 487. Stuttgart: Georg Thieme.

Gail A, Brinksmeyer HJ, Eckhorn R (1999b) Contributions of spike-rate and precise correlation to visual object representation in striate cortex of awake monkey. Soc Neurosci Abstracts 25:270.5.

Gail A, Brinksmeyer H-J, Eckhorn R (2000) Contour decouples gamma activity across texture representation in monkey striate cortex. Cereb Cortex 10:840-850.

Gail A, Brinksmeyer HJ, Eckhorn R (2001a) A dichoptic stimulator for simultaneous mapping of left and right eye cortical receptive fields during binocular vision with natural vergence in awake monkey. In: Proceedings of the 28th Göttingen Neurobiology Conference (Elsner N, Eysel U, eds), p 586, Stuttgart: Georg Thieme.

Gail A, Brinksmeyer HJ, Eckhorn R (2001b) Consistent perception-related modulations in monkey striate cortex during rivalrous and congruent dichoptic stimulation. Soc Neurosci Abstracts 27:123.6.

Gail A, Brinksmeyer HJ, Eckhorn R (2002a) Dichoptical RF-cinematogram – Simultaneous mapping of multiple binocular and both monocular receptive fields under natural vergence in awake animals. Submitted to J Neurosci Methods.

Gail A, Brinksmeyer H-J, Eckhorn R (2002b) Perception-related modulations of local field potential power and coherence in primary visual cortex of awake monkey during binocular rivalry. Submitted to Cereb Cortex.

Gegenfurtner KR, Hawken MJ (1996) Interaction of motion and color in the visual pathways. Trends Neurosci, 9:394-401.

Gilbert CD (1993) Circuitry, architecture, and functional dynamics of visual cortex. Cereb Cortex 3:373-386.

Gilbert CD, Das A, Ito M, Kapadia MK, Westheimer G (1996) Cortical dynamics and visual perception. Cold Spring Harbor Symposia on Quantitative Biology LXI, 105-113.

Glaser EM, Ruchkin DS (1976) Principles of neurobiological signal analysis. New York: Academic Press.

Goldstein EB (1989) Sensation and Perception. Third Edition. Belmont CA: Wadsworth.

Gray CM (1999) The temporal correlation hypothesis of visual feature integration: Still alive and well. Neuron 24:31-47.

Gray CM, König P, Engel AK, Singer W (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronisation which reflects global stimulus properties. Nature 338:334-337.

Grossberg S, Pessoa L (1998) Texture segregation, surface representation and figure–ground separation. Vision Res 38:2657–2684.

Güttler A, Eckhorn R, Frien A, Wölbern T (1996) Influence of micro-saccades on fast oscillations in monkey striate cortex. In: Proceedings of the 24th Göttingen Neurobiology Conference (Elsner N Eysel U, eds), 421. Stuttgart: Georg Thieme.

Güttler A, Eckhorn R, Jürgens E, Frien A (1997) Neural correlation contrast in visual cortex of monkey changes with stimulus contrast across an object-background-border. In: Proceedings of the 25th Göttingen Neurobiology Conference II (Elsner N, Waessle H, eds), p 551. Stuttgart: Georg Thieme.

Hebb DO (1949) The Organization of Behavior. New York: John Wiley & Sons.

Heeger DJ, Huk AC, Geisler WS, Albrecht DG (2000) Spikes versus BOLD: What does neuroimaging tell us about neuronal activity? Nat Neurosci 3:631-633.

Hegde J, van Essen D (2000) Selectivity for complex shapes in primate visual area V2. J Neurosci, 20 RC61:1-6.

Horn D, Sagi D, Usher M (1991) Segmentation, binding, and illusory conjunctions. Neural Computation 3: 510–525.

Hubel DH, Wiesel TN (1970) Stereoscopic vision in macaque monkey. Cells sensitive to binocular depth in area 18 of the Macaque monkey cortex. Nature 225:41-42.

Hubel DH, Wiesel TN (1977) Functional Architecture of Macaque Monkey Visual Cortex. Proc R Soc Lond, 198:1-59.

Hupé JM, James AC, Girard P, Lomber SG, Payne BR, Bullier J (2001) Feedback connections act on the early part of the responses in monkey visual cortex. J Neurophysiol 85:134-145.

Hupé JM, James AC, Payne BR, Lomber SG, Girard P, Bullier J (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. Nature 394:784-787.

Julesz B (1986) Texton gradients: The texton theory revisited. Biol Cybern, 54:245-251.

Jürgens E, Eckhorn R, Frien A, Wölbern T (1996) Restricted coupling range of fast oscillations in striate cortex of awake monkey. In: Proceedings of the 24th Göttingen Neurobiology Conference (Elsner N Eysel U, eds), 418. Stuttgart, New York: Georg Thieme.

Jürgens E, Güttler A, Eckhorn R (1999) Visual stimulation elicits locked and induced gamma oscillations in monkey intracortical- and EEG-potentials, but not in human EEG. Exp Brain Res 129:247-259.

Kandel ER, Schwartz JH, Jessell TM (1996) Neurowissenschaften. Heidelberg, Berlin, Oxford: Spektrum Akademischer Verlag.

Kapadia MK, Ito M, Gilbert CD, Westheimer G (1995) Improvements in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. Neuron 15:843-856.

Kapadia MK, Westheimer G, Gilbert CD (2000) Spatial distribution of contextual interactions in primary visual cortex and in visual perception. J Neurophysiol 84:2048-2062.

Kastner S, Nothdurft H-C, Pigarev IN (1997) Neural correlates of pop-out in cat striate cortex. Vision Res 37(4):371-376.

Kiper DC, Gegenfurtner KR, Movshon JA (1996) Cortical oscillatory responses do not affect visual segmentation. Vision Res, 36:539-544.

Knierim JJ, van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. J Neurophysiol 67(4):961-980.

Kobatake E, Tanaka K (1994) Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. J Neurophysiol, 71:856-867.

Koffka K (1935) Principles of Gestalt psychology. NewYork: Harcourt.

Kolb B, Whishaw IQ (1993) Neuropsychologie. Berlin, Heidelberg, Oxford: Spektrum Akademischer Verlag.

König P, Engel AK, Singer W (1995). Relation between oscillatory activity and long-range synchronization in cat visual cortex. Proc Natl Acad Sci USA 92:290–294.

König P, Engel AK, Singer W (1996) Integrator or coincidence detector? The role of the cortical neuron revisited. Trends Neurosci 19:130-137.

Kottmann M, Eckhorn R, Wölbern T, Frien A, Bauer R (1996) Synchronized fast oscillations in striate cortex correlate with perception of grating orientation in a binocular rivalry task in monkey. In: Proceedings of the 27th Göttingen Neurobiology Conference (Elsner N, Eysel U, eds), p 419, Stuttgart: Georg Thieme.

Kovács I, Julesz B (1993) A closed figure is much more than an incomplete one: Effect of closure in figure-ground segmentation. Proc Nat Acad Sci USA 90:7495-7497.

Kovács I, Papathomas TV, Yang M, Fehér Á (1996) When the brain changes its mind: Interocular grouping during binocular rivalry. Proc Natl Acad Sci USA 93:15508-15511.

Kreiter AK, Singer W (1992) Oscillatory neuronal responses in the visual cortex of the awake macaque monkey. Europ J Neurosci 4:369–375

Kreiter AK, Singer W (1996) Stimulus-dependent synchronization of neural responses in the visual cortex of the awake macaque monkey. J Neurosci 16:2381–2396.

Kremper A, Schanze T, Eckhorn R (2002) Classification of neural signals by a generalized correlation classifier based o radial basis functions. J Neurosci Methods 116:179-187.

Kruse W, Eckhorn R (1991) Inhibition of sustained gamma oscillations (35-80 Hz) by fast transient responses in cat visual cortex. Proc Nat Acad Sci USA 96:6112-6117.

Lamme VAF (1995) The neurophysiology of figure-ground segregation in primary visual cortex. J Neurosci 15:1605-1615.

Lamme VAF, Rodriguez-Rodriguez V, Spekreijse H (1999) Separate processing dynamics for texture elements, boundaries and surfaces in primary visual cortex of the macaque monkey. Cereb Cortex 9:406-413.

Lamme VAF, Roelfsema PR (2000) The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci 23:571-579.

Lamme VAF, Spekreijse H (1998) Neuronal synchrony does not represent texture segregation. Nature 396:362-366.

Lamme VAF, Supèr H, Landman R, Roelfsema PR, Spekreijse H (2000) The role of primary visual cortex (V1) in visual awareness. Vision Res 40:1507-1521.

Lamme VAF, Supèr H, Spekreijse H (1998) Feedforward, horizontal, and feedback processing in the visual cortex. Curr Opin Neurobiol :529-535.

Lamme VAF, Zipser K, Spekreijse H (1998) Figure–ground activity in primary visual cortex is suppressed by anesthesia. Proc Natl Acad Sci USA 95:3263-3268.

Lee S-H, Blake R (1999) Rival ideas about binocular rivalry. Vision Res 39:1447-1454.

Legatt AD, Arezzo J, Vaughan HG Jr. (1980) Averaged multiple unit activity as an estimate of phasic changes in local neuronal activity: effects of volume-conducted potentials. J Neurosci Methods 2:203-217.

Lehky SR (1988) An astable multivibrator model of binocular rivalry. Perception. 17:215-228.

Lehky SR, Maunsell JHR (1996) No binocular rivalry in the LGN of alert Macaque monkeys. Vision Res 36:1225-1234.

Leonards U, Singer W, Fahle M (1996) The influence of temporal phase differences on texture segmentation. Vision Res 36:2689-2697.

Leopold DA, Logothetis NK (1996) Activity changes in early visual cortex reflect monkey's percepts during binocular rivalry. Nature 379:549-553.

LeVay S, Connolly M, Houde J, van Essen DC (1985) The complete pattern of ocular dominance stripes in the striate cortex and visual field of the Macaque monkey. J Neurosci 5:486-501.

LeVay S, Voigt T (1988) Ocular dominance and disparity coding in cat visual cortex. Visual Neurosci 1:395-414.

Livingstone MS, Hubel DH (1984) Anatomy and physiology of a color system in the primate visual cortex. J Neurosci 4(1):309-356.

Livingstone MS, Hubel DH (1987) Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. J Neurosci 7:3416-3468.

Livingstone MS, Hubel DH (1988) Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. Science 240:740-749.

Livingstone MS, Tsao DY (1999) Receptive fields of disparity-selective neurons in macaque striate cortex. Nat Neurosci 2:825-832.

Logothetis NK (1998) Single units and conscious vision. Philos Trans R Soc Lond B Biol Sci 353:1801-1818.

Logothetis NK, Leopold DA, Sheinberg DL (1996) What is rivalling during rivalry? Nature 380:621-624.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150-157.

Lohmann H, Eckhorn R, Reitböck HJ (1988) Visual receptive fields of local intracortical potentials. J Neurosci Methods 25:29-44.

Luck SJ, Chelazzi L, Hillyard SA, Desimone (1997) Neural Mechanisms of Spatial Selective Attention in Areas V1, V2, and V4 of Macaque Visual Cortex. J Neurophysiol 77:24-42.

Lumer ED (1998) A neural model of binocular integration and rivalry based on the coordination of action-potential timing in primary visual cortex. Cereb Cortex 8:553-561.

Lumer ED, Friston KJ, Rees G (1998) Neural correlates of perceptual rivalry in the human brain. Science 280:1930-1934.

Macy A, Ohzawa I, Freeman RD (1982) A quantitative study of the classification and stability of ocular dominance in the cat's visual cortex. Experimental Brain Research, 48:401-408.

Marmarelis PZ, Marmarelis VZ (1978) Analysis of physiological systems.New York: Plenum Press.

Maunsell JHR, Newsome WT (1987) Visual processing in monkey extrastriate cortex. Annu Rev Neurosci 10:363-401.

Mersereau RM, Speake TC (1983) The processing of periodically sampled multidimensional signals. IEEE Trans Acoust Speech Signal Process 31:188-194.

Milner PM (1974) A model for visual shape recognition. Psychol Rev 81:521-535.

Mitzdorf U (1985) Current source-density method and application in cat cerebral cortex: Investigation of evoked potentials and EEG phenomena. Physiol Rev 65(1):37–100.

Mitzdorf U (1987) Properties of the evoked potential generators: Current source-density analysis of visually evoked potentials in the cat cortex. J Neurosci 33:33-59.

Ngo TT, Miller SM, Liu GB, Pettigrew JD (2000) Binocular rivalry and perceptual coherence. Curr Biol 10:134-136.

Nicholls JG, Martin AR, Wallace BG (1992) From Neuron to Brain. Third Edition. Sunderland: Sinauer Associates, Inc.

Niessing M (1999) Mikroaugenbewegungen und ihr Einfluß auf Signale im Sehcortex. Diploma thesis. Department of physics, Philipps University, Marburg.

Nowak LG, Bullier J (1997) The timing of information transfer in the visual system. In: Cerebral Cortex 12. Extrastriate cortex in primates. Rockland et al. (eds.). Plenum Press, New York.

Olshausen BA, Anderson CH, Van Essen DC (1993) A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. J Neurosci 13:4700-4719.

Olshausen BA, Anderson CH, Van Essen DC (1995) A multiscale dynamic routing circuit for forming size- and position-invariant object representations. J Comp Neurosci 2:45-62.

Palm G (1982) Neural assemblies. An alternative approach to artificial intelligence. Berlin Heidelberg NewYork: Springer.

Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes. Biophys J 7:420–440

Perrett DI, Rolls ET, Caan W (1982) Visual neurons responsive to faces in the monkey temporal cortex. Exp Brain Res 47:329-342.

Poggio GF (1995) Mechanisms of stereopsis in monkey visual cortex. Cereb Cortex 3:193-204.

Poggio GF, Gonzalez F, Krause F (1988) Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity. J Neurosci 8:4531-4550.

Polat U, Sagi D (1993) Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. Vision Res 33:993-999.

Polat U, Sagi D (1994) The architecture of perceptual spatial interactions. Vision Res 34:73-78.

Pollen DA (1999) On the neural correlates of visual perception. Cereb Cortex 9:4-19.

Polonsky A , Blake R, Braun J, Heeger DJ (2000) Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. Nat Neurosci 3:1153-1159.

Prince SJD, Cumming BG, Parker AJ (2002a) Range and mechanism of encoding of horizontal disparity in Macaque V1. J Neurophysiol 87:209-221.

Prince SJD, Pointon AD, Cumming BG, Parker AJ (2002b) Quantitative analysis of the responses of V1 neurons to horizontal disparity in dynamic random-dot stereograms. J Neurophysiol 87:191-208.

Rees G, Friston K, Koch C (2000) A direct quantitative relationship between the functional properties of human and Macaque V5. Nat Neurosci 3:716-723.

Rees G, Kreiman G, Koch C (2002) Neural Correlates of Consciousness in Humans. Nat Neurosci Rev 3:261-270.

Reid RC, Victor JD, Shapley RM (1997) The Use of m-sequences in the analysis of visual neurons: Linear receptive field properties. Visual Neurosci 14:1015-1027.

Reitböck HJ (1983) A multi-electrode matrix for studies of temporal signal correlations within neural assemblies. In: Synergetics of the brain (Basar E, Flohr H, Haken H, Mandell AJ, eds), p174-182. Berlin, Heidelberg, New York: Springer.

Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. J Neurosci 19:1736-1753.

Reynolds JH, Desimone R (1999) The role of neural mechanisms of attention in solving the binding problem. Neuron 24:19-29.

Rockland, KS, van Hoesen GW (1994) Direct temporal-occipital feedback connection to striate cortex (V1) in the macaque monkey. Cereb Cortex 4:300-313.

Rockland KS, Drash GW (1996) Collateralized divergent feedback connections that target multiple cortical areas. J Comp Neurol 373:529–548.

Rockland KS, Lund JS (1983) Intrinsic laminar lattice connections in primate visual cortex. J Comp Neurol 216:303-18

Rockland KS, Pandya DN (1979) Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. Brain Res 179:3-20.

Roelfsema PR, Engel AK, König P, Singer W (1996) The role of neuronal synchronization in response selection: A biological plausible theory of structured representations in the visual cortex. J Cogn Neurosci 8(6):603-625.

Roelfsema PR, Lamme VAF, Spekreijse H (1998) Object-based attention in the primary visual cortex of the macaque monkey. Nature 395:376-381.

Roelfsema PR, Lamme VAF, Spekreijse H, Bosch H (2002) Figure-ground segregation in a recurrent network architecture. J Cogn Neurosci 14:525-537.

Saam M, Eckhorn R (2000) Lateral spike conduction velocity in the visual cortex affects spatial range of synchronization and receptive field size without visual experience: A learning model with spiking neurons. Biol Cybern 83:L1-L9.

Saam M, Gabriel A, Al-Shaikhli B, Eckorn B (2000) Neural mechanisms of figure-ground segregation. Recordings from awake monkey and a model with spiking neurons. In: Proceedings in artificial intelligence 9. Dynamische Perzeption (Baratoff G, Neumann H, eds), pp 27-32. Berlin: Infix.

Salin PA, Bullier J (1995) Corticocortical connections in the visual system: Structure and function. Physiol Rev 75:107-154.

Schanze T, Eckhorn R (1997) Phase correlation among rhythms present at different frequencies: Spectral methods, application to microelectrode recordings from visual cortex, and functional implications. Int J Psychophysiol 26:171-189.

Schröder JH (1996) Analysemethoden für die raumzeitliche Dynamik rezeptiver Felder des visuellen Kortex von wachen Affen. Diploma thesis. Department of physics, Philipps University, Marburg.

Sheinberg DL, Logothetis NK (1997) The role of temporal cortical areas in perceptual organization. Proc Natl Acad Sci USA 94:3408-3413.

Singer W (1999) Neuronal Synchrony: A versatile code for the definition of relations? Neuron, 24:49-65.

Srinivasan R, Russell DP, Edelman GM, Tononi G (1999) Increased synchronization of neuromagnetic responses during conscious perception. J Neurosci 19:5435-5448.

Stöcker M, Reitböck HJ, Eckhorn R (1996) A neural network for scene segmentation by temporal coding. Neurocomputing, 11:123–134.

Stoerig P (1996) Varieties of vision: from blind responses to conscious recognition. Trends Neurosci 19:401-406.

Stoerig P (2001) The neuroanatomy of phenomenal vision: a psychological perspective. Ann N Y Acad Sci 929:176-94.

Stoerig P, Cowey A (1997) Blindsight in man and monkey. Brain 120:535-59.

Stoerig P, Zontanou A, Cowey A (2002) Aware or unaware: Assessment of cortical blindness in four men and a monkey. Cereb Cortex 12:565-574

Sutter EE (1992) A deterministic approach to nonlinear systems analysis. In: Nonlinear vision: determination of neural receptive fields, function, and networks (Pinter RB, and Nabet B, eds), pp 171-220. Boca Raton: CRC Press.

Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci 3:151-162.

Tallon-Baudry C, Bertrand O, Fischer C (2001) Oscillatory synchrony between human extrastriate areas during visual short-term memory maintenance. J Neurosci 21(RC177):1-5.

Tanaka K (1996) Inferotemporal cortex and object vision. Annu Rev Neurosci 19:109-139.

Thorpe S, Fize D, Marlot C (1996) Speed of processing in the human visual system. Nature 381:520-522.

Tong F, Engel SA (2001) Interocular rivalry revealed in the human cortical blind-spot representation. Nature 411:195-199.

Tong F, Nakayama K, Vaughan JT, Kanwisher N (1998) Binocular rivalry and visual awareness in human extrastriate cortex. Neuron 21:753-759.

Tononi G, Srinivasan R, Russell PD, Edelman GM (1998) Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses. Proc Natl Acad Sci USA 95:3198-3203.

Treisman A (1996) The binding problem. Curr Opin Neurobiol 6:171-178.

Treisman A (1998) Feature binding, attention and object perception. Philos Trans R Soc Lond B Biol Sci 353:1295-1306.

Treisman A, Gelade G (1980) A feature-integration theory of attention. Cogn Psychol 12:97-136.

Treisman A, Sato S (1990) Conjunction search revisited. J Exp Psychol 16:459-478.

Treisman A, Schmidt H (1982) Illusory conjunctions in the perception of objects. Cogn Psychol 14:107-141.

Trotter Y, Celebrini S, Stricanne B, Thorpe S, Imbert M (1996) Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. J Neurophysiol 76:2872-2885.

Ungerleider LG and Mishkin M (1982) Two cortical visual systems. In: Analysis of visual behavior (Ingle DJ, Goodale MA, Mansfield RJW, eds), p549-586.Cambrigde MA: MIT press.

Vidyasagar TR (1998) Gating of neuronal responses in macaque primary visual cortex by an attentional spotlight. NeuroReport 9:1947-1952.

Vijn PCM, van Dijk BW, Spekreijse H (1991) Visual stimulation reduces EEG activity in man. Brain Res 550:49–53.

von der Malsburg C, Schneider W (1986) A neural cocktail-party processor. Biol Cybern 54:29-40.

von Stein A, Chiang C, König P (2000) Top-down processing mediated by interareal synchronization. Proc Natl Acad Sci USA 97:14748-14753.

von Stein A, Sarnthein J (2000) Different frequencies for different scales of cortical integration: From local gamma to long range alpha/theta synchronization. Int J Psychophysiol 38:301-313.

Wertheimer M (1923) Untersuchungen zur Lehre von der Gestalt: II. Psychologische Forschung 4:301-350.

Wölbern T, Eckhorn R (2002) Perceptual grouping correlates with short synchronization in prestriate cortex of a monkey. Neuroreport (in press).

Wolfe JM (1983) Influence of spatial frequency, luminance, and duration on binocular rivalry and abnormal fusion of briefly presented dichoptic stimuli. Perception 12:447-456.

Wolfe JM, Cave KR (1999) The Psychophysical Evidence for a Binding Problem in Human Vision. Neuron 24:11-17.

Wolfe JM, Cave KR, Franzel SL (1989) Guided search: An alternative to the feature integration model for visual search. J Exp Psychol 15:419-433.

Yoshioka T, Blasdel GG, Levitt JB, Lund JS (1996) Relation between patterns of intrinsic lateral connectivity, ocular dominance, and cytochrome oxidase-reactive regions in Macaque monkey striate cortex. Cereb Cortex 6:297-310.

Zeki S, Bartels A (1999) Toward a theory of visual consciousness. Conscious Cogn 8:225-259.

Zhou H, Friedman HS, von der Heydt R (2000) Coding of visual border ownership in monkey visual cortex. J Neurosci 20:6594-6611.

Zipser K, Lamme VAF, Schiller PH (1996) Contextual modulation in primary visual cortex. J Neurosci 16:7376-7389.

Abbreviations

BOLD   blood oxygenation level dependent

BR   binocular rivalry

CMR   co-modulation ratio

CRF   classical receptive field

DCRF   dichoptical classical receptive field (mapping)

EEG   electroencephalogram

fMRI   functional magnetic resonance imaging

LFP   local field potential

LGN(d)   lateral geniculate nucleus (dorsal)

MEG   magnetoencephalogram

MUA   multiple unit acitvity

ODI   ocular dominance index

PSTH   peri-stimulus time histogram

ROI   region of interest

RF   receptive field

SD   standard deviation

SNR   signal-to-noise ratio

SUA   single unit activity

TOI   time of interest

V1   primary visual cortex (also: striate cortex)

Glossary

Action potential Fast, transient, impulse-like (~1 ms; ~100 mV) electrical membrane depolarisation actively generated in an all-or-nothing manner and transmitted along the axon of a neuron*.

Binocular rivalry Perceptual phenomenon induced by incongruent visual stimulation of the left and right eye with non-fuseable patterns. Subjective perception stochastically switches between the two alternatives offered via the two eyes.

Coherence Linear, spectral selective coupling measure for pairwise signals, here exclusively used for time series (for a formal definition cf. Paragraph 4.2.2). At each frequency coherence is independently normalized, and is sensitive to covariation of amplitude densities and constancy of relative phase at this frequency.

Figure–ground segregation Segregation of a visual pattern into an object (figure) seen in front of a background; the background is assumed to extend behind the object.

Local field potential, LFP Local superposition of slow somato-dendritic potentials* of a group of neurons*. Captured by extracellular microelectrode recordings after low-pass filtering (1–140 Hz; 18 db/oct). Estimated radius of support (half-height decline): ~500 μm.

Multiple unit activity, MUA Local action potential (spike)* density of a group of neurons*. Captured by extracellular microelectrode recordings after band-pass filtering (1–10 kHz; 18 db/oct), full-wave rectification, and subsequent low-pass filtering (140 Hz; 18 db/oct). Estimated radius of support (half-height decline): ~50 μm.

Neuron Main cell type in the nervous system (~10^{11} [10^5/mm^3] in the human brain); comprising several classes, e.g., pyramidal cells, basket cells, stellar cells, etc. A typical central neuron consists of three main parts: dendrites, soma, and axon. **Dendrites** form an often wide-spread arborized, receptive region of a neuron integrating synaptic inputs (of typically ~10^3–10^4 synapses* in a central neuron). **Soma** refers to the cell body, comprising metabolistic structures and a specialised spike-endcoding region. The **axon** is the long-tailed, action potential* carrying output structure of a neuron, starting from a specialized spike-encoding region of the soma, eventually diverging into many collaterals, and terminating in presynaptical endings at other neurons' dendrites, somata, or axons (or effectors outside the central nervous system).

Peri-stimulus time histogram, PSTH   Neuronal response time-signals averaged over repetitions of identical stimulus presentations and triggered with respect to the points in time of the stimulus onsets.

Primary visual cortex, V1   First cortical area to receive input via the main visual sensory pathway. In primates identical to the anatomical-histological definition of Brodman area 17.

Receptive field, visual   The classical receptive field concept refers to the spatial area of photoreceptors in the retina activating a certain neuron when elicited. The term is synonymously used for the corresponding area of the visual field, and is mostly defined as a *minimum response field* (Barlow, 1972) determined with a small test stimulus. Receptive fields are typically characterized by extra stimulus specificities, e.g., for colour, orientation of lokal contrast borders, direction of motion, etc. The receptive field concept can also be extended by its dynamic properties, or by the definition of additional regions of modulatory influence coming into play only if a neuron is simultaneously activated via stimulation in the classical receptive field.

Single unit activity, SUA   Action potential (spike)* output of a single neuron*.

Somato–dendritic potentials   Postsynaptic, graded, de- or hyper-polarising membrane potential changes at the dendrites or soma of a neuron*; *slow* potentials compared to axonal action potentials* due to longer time constants of the somato–dendritic membrane (typically by a factor of ~10–100).

Spike   cf. action potential

Striate cortex   cf. primary visual cortex

Synapse   Connecting structure between neurons* mediating electrical or chemical signal transmission ($\sim 10^{14}$ [10^8/mm^3] chemical synapses in the human brain).

Visual awareness   Subjective conscious experience of a visual sensory stimulus. Unless otherwise indicated the term *perception*, as used here, includes the state of visual awareness.

Scientific CV

Name:	Alexander M. Gail
Date of birth:	17th October 1970
Place of birth:	Augsburg, Germany
Nationality:	German
Present institution:	Neurophysics Group, Dept. Physics,
	Philipps University Marburg.
	Renthof 7, 35032 Marburg, Germany
	Tel.: ++49 (0) 6421 / 282-66 31
	Email: alexander.gail@physik.uni-marburg.de

EDUCATION:

since 11/1997 — **Research Associate**
in DFG research group "Dynamics of Cognitive Representation",
Neurophysics Group, Dept. Physics,
Philipps University Marburg

02/2000 — **Course on Laboratory Animal Science**
School of Veterinary Medicine Hannover

07/1997 — **Physics Diploma**
Subjects of examination: Theoretical Physics, Experimental Physics, System Theory and System-Neurophysiology, Cognitive Psychology and Neuropsychology;
Diploma thesis: "Signal correlations in the visual cortex of awake monkey in the context of learning processes".

10/1993 – 07/1997 — **Physics studies** (continued)
at Philipps University Marburg.

10/1991 – 09/1993 — **Physics studies** including Vordiplom (intermediate examination)
at University of Augsburg

06/1990 — **Abitur**
Wernher-von-Braun Gymnasium, Friedberg/Bay.