# STUDYING FEATURE SPECIFIC MECHANISMS OF THE HUMAN VISUAL SYSTEM

**Christian Emanuel Kaul**

Institute of Cognitive Neuroscience
Wellcome Department of Imaging Neuroscience
Institute of Neurology
University College London

Prepared under the supervision of:
Professor Geraint Rees
Professor Nilli Lavie
Submitted to UCL for the Degree of PhD

**Declaration**

I, Christian Kaul, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in my thesis.

Part of the work presented in Chapter VI has been published as the following paper:

Kaul C, Bahrami B (2008) Subjective experience of motion or attentional selection of a categorical decision. *Journal of Neuroscience* 28:4110-4112.

**Acknowledgements**

This work has been funded by a BRAIN research trust prize and Geraint's Wellcome Trust Senior Fellowship.

I would like to thank all the past and present people in my lab here in London: Geraint, Bahador, Elaine, Sam, Frank, Claire, Ayse, Marieke, Rimona, Lauri, Richard, David, Su, Julia, Elisa, Jen, Claire, Nahid, Vik and Nilli.

Thanks also go to Alumit, Julio and Philipp who welcomed me in their labs for scientific visits.

Special thanks go to Bahador whom I often went first whenever an issue needed discussing – he always had time, an open ear and most of the time: good advice. Significant teaching and help I also received from Richard and Elaine. Thank you all so much.

My time in London was one of great fun, excitement and excellent mood. Much of this is due to the support I have been given by all of the above, but none more than Geraint. From the first day, Geraint constantly supported me, helped me through the good and the hard times. He always believed in me which was the number one most motivating thing in my entire PhD time. Together we achieved a great deal of personal and professional achievements and I am proud and honored to having had the chance to do my PhD with him. I will never be able to thank you enough! For the future I truly hope we remain the good friends that we are now.

Finally, I would like to thank my family and friends and especially Tina – you have been the light of my life for the last 15 months and a great deal of my smooth PhD-experience is thanks to you.

**Abstract**

What are the current limits of our knowledge of brain activity underlying vision and can I further this knowledge? In this thesis, I explore this basic question. I focus on those aspects of visual input that can be described as basic features of visual perception. Examples include orientation, color, direction of motion and spatial frequency. However, understanding how humans visually perceive the external world is closely related with the study of attention. Attention, that is, the selection of some aspects of the environment over others, is one of the most intensively studied areas in experimental psychology, yet its neural mechanisms remain largely elusive.

This thesis focuses on three distinct topics at the border of feature-specific visual perception and feature-specific visual attention. First, in a series of studies, I explore the influence of heightened attentional demand to a central task to feature-specific neural processing in the ignored periphery. I discover that heightened attentional demand does not influence feature-specific representations in early visual cortices. Second, I investigate the influence of feature-based attention on neural processing of early visual cortices. At the same time, I also probe the influence of a behavioral decision to deploy feature-specific attention in the imminent future. I find that feature-based attention operates independent of other types of attention. Additionally, results indicate that a behavioral decision to deploy feature-based attention alone, without visual stimulation present, is able to modulate neural

activity in early visual cortices. Third, I examine the more complex feature of facial gender and where in the brain gender discrimination might receive neural processing. I find that, in an established network of face-selective brain areas, facial gender is represented in nearly all areas of that network. Finally, I discuss all findings in the light of the current state of research, for their scientific significance and for future research opportunities.

**Table of Contents**

## Preface

Vision has been a central topic of study in neuroscience for over a hundred years. In particular, the relationship between attention and visual processing has been studied extensively. Attention modulates the level of brain activity and behavioral performance associated with the processing of visual stimuli. In humans many studies have addressed the neural basis of attentional modulation of visual perception using non-invasive techniques, like electroencephalography (EEG) and functional magnetic resonance imaging (fMRI). However, it has proven difficult to make clear inferences about the effect of attention on underlying neural representations of visual stimuli. Difficulties arise mostly due to methodological restrictions like the limits of spatial and temporal resolution of non invasive techniques and the complex task of inferring underlying neural activity from results obtained by these methods. As a consequence, it remains largely elusive what lies within the general observation that activity in a neuronal population associated with visual processing is modulated, often enhanced, by visual attention. However, it is the nature of this modulation of the underlying neuronal representation of visual stimuli, which has fascinated the imagination of many visual neuroscientists.

In this thesis, I explore basic questions about visual processing, in a variety of paradigms and settings, but always with a central question in mind: What are the current limits of our knowledge of brain activity underlying vision and can I further this knowledge? Naturally, I limit myself to very specific sub-aspects of visual

processing, namely the representation of irrelevant basic visual features (like orientation, hue, color or direction of motion) under differing attentional load conditions, the modulation of visual processing by feature based attention and the representation of facial gender information in the brain.

In the next sections I review the research that prompted the questions addressed in this thesis. I begin with a brief description on the debate regarding the selection process of visual attention, followed by a description of load theory, which offers a resolution to the debate. I follow this with a brief, non-technical, review of multivariate pattern recognition for fMRI and why this new methodological approach is highly relevant when studying brain activity underlying feature-specific processes in the human brain. This is followed by an outline of the evidence for different types of attention, with particular focus to the particular instance of feature-based attention and feature-specific selection. Finally, I describe in brief the current state of knowledge on face-specific visual processing and how it relates to this thesis. I close with an outlook on the specific questions addressed in studies of this thesis. In a separate chapter I review, in detail, methods and techniques used throughout the experimental work described in this thesis.

# Chapter I:

# General Introduction

## I.1 Load Theory

### I.1.1 Visual attention

What humans perceive at any given moment is not rigidly determined by the sensory inputs available. Instead, humans have the ability to select a subset of the available perceptual information. In the study of vision, this active process of visual perception has been labeled *visual attention* (Treisman, 2004). Before visual processing was widely examined at the level of brain activity, in the early days of attention research, visual information processing was often seen as a simple pipeline of successive stages: information travelling from one stage to another. In this simplified model visual attention was commonly described as a filter, acting at one of these different stages to select information for further processing while excluding irrelevant or unattended information. With evidence from neuroscience this pipeline view was, however, soon replaced with a more complex system allowing for parallel pathways, top-down feedback and lateral connections. Of course the new, more complex model, allowed for visual attention to act more flexibly. None the less, for a long time, one central discussion in studies about visual attention remained whether any filtering or selection occurred

"early" or "late" in visual perception. This was of course confusing in the first place, as to answer the question whether attention acts 'early' or 'late' depends on how to interpret the words early and late: for example, as measures of time, of processing order or of neuroanatomy. At least partly as a consequence of this confusion, most research focused on answering a slightly more specific question: whether visual perception is an automatic process. In one popular view this automatic processing was argued to occur up to a very high level of detail (i.e. object recognition) and attentional selection was mainly described as acting 'late' - at the level of memory and response selection (Deutsch and Deutsch, 1963). In contrast, another popular view held that attentional selection acted 'early' - at the level of perception itself (Broadbent, 1958). This discussion was not trivial as both views were backed up with many apparently competing behavioral experimental findings. However, at the time, the actual concepts discussed were mainly of a theoretical nature, describing psychological models rather than underlying physiological processes. Over the years and fueled by numerous experimental results providing equally convincing evidence for early and late selection, often contradicting each other, the discussion seemed unsolvable. In fact, as recent as 1993, it was suggested that the contradictions thrown up by different research studies about visual attention might never be resolved (Allport, 1993)

### I.1.2 Does attention act early or late in visual processing

Why is it reasonable to think of visual attention as a process acting early in visual processing? As an example consider only the earliest stage of visual information processing: at any given time, only a fraction of the information received from the retina can be selected for further processing. Yet, humans have to ability to select a subset of the available perceptual information. Thus, two fundamental properties of the visual perception system were generally listed as relevant for early selection: first, visual perception has limited capacity for processing information, and second, visual perception is selective (Broadbent, 1958). A common theoretical interpretation emerged: it must be that visual attention acted as a filter early in processing of visual stimuli. This led to the early selection view, supported by many studies (Broadbent, 1958; Rock and Gutman, 1981; Mack, 1998). Early selection models propose a number of perceptual inputs competing for a central selective filter. From all competing inputs a small subset of inputs are selected by this filter for further processing (i.e. "attending to"). Selected information is passed along to higher areas. Thus, as a result, unattended information is filtered out early and does not influence further processing.

In contrast, according to the late selection view, visual perception occurs automatic and preattentively. The crucial difference to the early selection studies is that unattended stimuli are often shown to influence perception. Thus instead of visual signals being filtered out early, visual perception is argued to occur automatically for all

stimuli, making all information available. Then visual attention selects a subset of already processed information for further analysis, response planning, memory tasks, etc (Deutsch and Deutsch, 1963; Treisman and Geffen, 1968; Neill, 1977; Tipper, 1985). Hence, late selection views postulate that perception is an automatic process, which proceeds on all stimuli regardless of their task relevance. Attention, according to this view, can only affect post-perceptual processing stages such as visual selection along guidelines from higher areas.

The debate has proved to be of surprising longevity, mostly due to the fact that substantial empirical support has been found for both theories. Reviewing the literature in 2001, Jon Driver concluded that both views together resulted in "a rather confusing picture of visual attention for a long time" (from: A selective review of selective attention research of the past century, Driver, 2001).

### I.1.3 Load theory

Lavie (1995) proposed a theory to elegantly solve this long-standing debate of early versus late selection. Her original theory was later expanded to the "Load theory of selective attention and cognitive control" (Lavie, 1995; Lavie et al., 2004). In this theory attentional resources are allocated involuntarily to process relevant and irrelevant stimuli up to a capacity limit, thus the theory combines aspects of both the early and late selection viewpoints and accounts for the contradictory results found in earlier research. Consequently,

Lavie's load theory allows a reinterpretation of previous experimental work and gives rise to novel, empirically-testable predictions. The main components of Lavie's theory are attentional load (also named perceptual load) and cognitive control. In this thesis the terms attentional load and perceptual load are used interchangeably.

Attentional load is a mechanism that primarily deals with the amount of visual information at a given time in an attended situation. Attentional selection occurs up to the capacity limits of visual attention, thus, the perceptual system does indeed have limited capacity (as proposed by early selection). However, Lavie's theory also proposes that all stimuli, regardless of their relevance to the task at hand, are processed automatically (as in late selection) – but only until perceptual capacity is exhausted (Lavie and Tsal, 1994; Lavie, 1995). This results in the interesting proposal that low attentional load during an attended goal-directed task leaves spare attentional capacity to 'spill over' to goal-irrelevant features. On the other hand, high attentional load takes up all capacity and thus decreases distractor processing. Attentional load finds its counterpart in executive control: an active attentional control mechanism determining stimulus processing priorities, e.g. between targets and irrelevant distractors. Cognitive control helps to keep track of changes, goals or intentions and is acting within the range of working memory. The cognitive control mechanism keeps priority of attended stimuli or features and by doing so reduces interference from distractors. Working memory is one cognitive mechanism to maintain such prioritization. Importantly, cognitive control under

high working memory load should have an opposite effect to that obtained under high perceptual load. An increased condition leaves no spare working memory capacity to reduce interference (maintain prioritization) and will therefore lead to increased distractor processing. Work in this thesis exclusively focuses on the mechanisms of attentional load.

### I.1.4 Evidence for attentional load as proposed in the load theory

*Behavioral studies of attentional load.* Load theory maintains that the attentional load imposed by a task determines the extent of distractor processing. A review of previous research showed that evidence of 'early selection' was usually found in studies in which the task involved considerable attentional load, leading to an exhaustion of perceptual capacity and therefore reduced distractor interference (Rock and Gutman, 1981; Lavie and Tsal, 1994). On the contrary, in studies consisting of low load target(s) and distractor(s) there generally was enough spare capacity for the irrelevant distractor to be perceived and processed. Thus, in these studies visual attention was often found to act late, and irrelevant, unattended, distractors often influenced current perception (i.e. Hagenaar and van der Heijden, 1986; Lavie and Tsal, 1994).

However, since attentional load was not directly manipulated in any of the previous studies, it was still conceivable that the discrepancies in findings could be due to alternative factors.

Consequently, load theory has been tested empirically by Lavie herself and others. Numerous experimental paradigms have empirically tested the influence of attentional resources to unattended stimuli in the periphery (distractor) while subjects performed a central task of varying attentional load. Experiments commonly deprive visual distractor-stimuli of attention by manipulating attentional load in an unrelated task. The level and type of load of a goal-directed task drastically influences distractor processing in humans. High attentional load can severely reduce distractor processing (Lavie, 1995; Lavie and Fox, 2000; see Lavie et al., 2004 for full theory and many more examples).

Later on, these insights were exploited to produce most striking demonstrations of highly demanding attentional tasks exhausting visual perceptual capacity so effectively, that even most salient distractors become invisible under high load, but are visible under low load (Cartwright-Finch and Lavie, 2007). In another example (for high load only), Simons & Chabris (1999) showed participants a 25 second video clip, in which two teams – one wearing white shirts, the other black – passed balls between members of the same team. While participants monitored the white team, counting passes, a person in a black gorilla suit walked across the screen. When asked at the end of the clip, participants often failed to report the gorilla even though it is in plain sight for 9 seconds (Simons and Chabris, 1999). While the study of Simons & Chabris (1999) lacks a low load condition to compare, it still remains one of the most striking demonstrations of the influence of high load on conscious perception.

An example of one experimental paradigm that demonstrates the effects of attentional load on irrelevant distractor processing is shown in Figure I-1 (Lavie and Cox, 1997; see also Lavie, 2005). In the task participants made speeded responses indicating whether a target letter in a ring of letters was one of two pre-specified letters (X or N) while attempting to ignore distractor letters in the ring and in the periphery. Under low attentional load all letters of the central circle were simple circles. The presence of a congruent distractor (the same letter than the target) outside the circle significantly increased reaction times as compared to the presence of an incongruent distractor (Figure I-1 B, red bar). Thus, the authors concluded that the easy task left spare attentional capacity used to perceive the distractor, resulting in behavioral interference with the target identification. However, when the task was transformed into a high load version, by replacing all symbols of the circle with random letters, the task became significantly more difficult. Now, under high attentional load, the effect of congruent vs. incongruent distractor was not significant anymore (Figure I-1 B, green bar). Thus, the authors concluded that, under high load, the distractor letter was not processed anymore due to insufficient processing capacity.

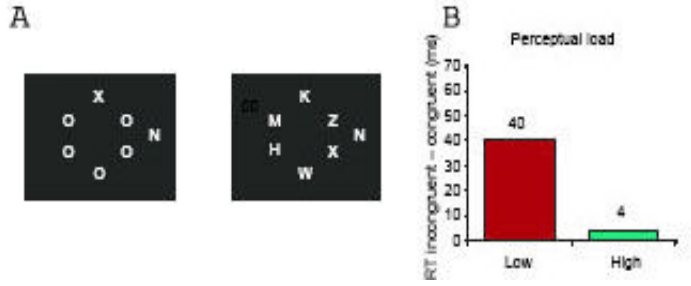



**Figure I-1**
**A – Typical example of an attentional load experiment. Subjects make speeded responses indicating whether a central target letter in one of two pre-specified letters (X or N) while attempting to ignore a peripheral distractor letter.**
**B - Reaction time difference between different attentional load conditions.**

Following these early examples of the modulatory role of attentional load, there is an ever growing body of studies showing that the processing of irrelevant distractors is reduced under high attentional load. Experiments span a wide variety from more classical effects on distractors consisting of color-shape conjunctions to working memory related measures (Lavie, 1997; Lavie and Fox, 2000; Jenkins et al., 2003, 2005). More recently high load was additionally shown to reduce effects of various other measures, including individual differences in distractibility and, finally, relevant distractors (Forster and Lavie, 2007, 2008).

*Functional imaging studies of attentional load.* While the load theory provided an explanation for many behavioral results, behavioral studies of visual attention tasks do not establish the underlying neural mechanisms. Thus, soon after the load theory of attention was proposed, studies of the underlying neural correlates of attentional load were carried out. Recent research has produced many insights about the influence on the guidance of visual

attention and its relationship to visual awareness (Rees and Lavie, 2001). Other than behavioral results, functional imaging studies have the advantage that the effect of load can be (indirectly) measured on brain activity (as measured by the BOLD signal) of a distractor rather than a behavioral measure related to a distractor. Thus, experiments commonly deprive visual stimuli of attention by manipulating attentional load in an unrelated task to then measure the influence of this load manipulation on the distractor brain activity. One of the earliest demonstrations of changes on distractor processing caused by differing attentional load in an unrelated task, the so called "load effect", was the finding of reduced activity associated with motion processing in human V5/MT when comparing task-irrelevant moving vs. static distractor dots under high vs. low load in a central task (Rees et al., 1997). Since then several such studies established the effect of high attentional load in different areas of the brain, reaching from the sub cortical level of the lateral geniculate nucleus (O'Connor et al., 2002), amygdala (Pessoa et al., 2002) to striate cortex (Schwartz et al., 2005) and higher brain areas such as the parahippocampal place area (Yi and Chun, 2005) and V4 (Pinsk et al., 2004). The stimuli used vary from meaningless checkerboards to meaningful pictures and even meaningful facial expressions.

Taken together, all these studies show that functionally specialized regions of visual cortex are less activated by irrelevant distractors under conditions of high attentional load compared to low load conditions, consequently implying that processing of irrelevant visual distractors depends on availability of attention.

### I.1.5 New directions in Load theory research

While many studies demonstrate a striking effect of attentional load on overall brain activity levels, most do not clarify exactly what aspect of neuronal activity related to the fMRI signal is altered. This is especially true for the neural representation of non-target related stimuli (distractors). Thus, an important open question is which aspect of neuronal activity is modulated by load. For example, considering load modulations on distractor activity in early visual cortex, it remains unclear whether the reduced activity in V1 points towards the possibility that the distractor was (partly) less processed or less attended or both (Schwartz et al., 2005). Other non-exclusive alternatives could be that distractor representations in that experiment received less feedback from higher areas under high load, or were less synchronized with other brain areas. The question of what the observation of a reduction in fMRI signal within a single area indicates in terms of the underlying neuronal representations and their coupling with other areas is not easily answered, since it is unclear what kind of data allow deeper insights into the information content of neural activity in V1. In the example of Schwarz et al (2005), the activity modulated by attentional load arose from stimulation with flickering checkerboards. In this example, how is the neural representation of the checkerboard-distractor altered beyond the finding of decreased activity under high load? Are the high-contrast borders of the checkerboard represented to a lesser degree? Or did participants process certain

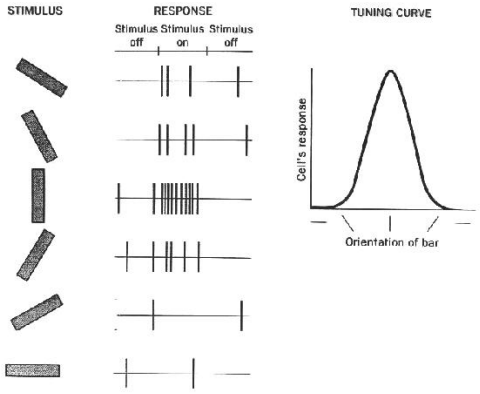
aspects of the checkerboard less under high load than under low load? To answer these questions is non-trivial as task requirements are such that the stimulus remains an irrelevant distractor, thus, direct behavioral measures are hard to obtain. One potential way to address this problem is to design the distractor in a way that is particularly well understood in the terms of its neural architecture and representation. For the brain areas in question, early visual cortices, it is well known that activity of neurons are tuned to elementary features such as color and orientation (Hubel and Wiesel, 1962, 1968). In fact, the columnar architecture in primary visual cortex that represents specific features of the visual input is one of the best studied neuronal architectures in the brain (e.g. colour, orientation or direction of motion, Mountcastle, 1997). Further, while in early visual cortex, neuronal responses reflect elementary features (such as the orientation of edges) in the visual environment, later in visual processing, visually responsive neurons show selectivity for particular categories of visual stimuli (e.g. faces). While the spatial representation in higher areas is mostly unknown, basic visual features are commonly represented at a spatial scale of a few hundred microns, with different locations in visual cortex corresponding to different preferences for such elementary features (see chapter II for more details). For the purpose of attentional load research and the question of how the neural representation of irrelevant distractors is modulated by different attention load conditions, it might thus prove fruitful to investigate the effect of load on such a representation of a basic visual feature in the brain. However, such representations have been thought inaccessible to non-invasive imaging techniques in

humans. This is because techniques such as fMRI or EEG in humans typically record at spatial resolutions far greater than the size of such columns of brain tissue representing different visual stimulus features. Hence, differences in elementary features were thought to be not measurable with functional brain imaging. As a consequence, there is a gap in our knowledge about the exact qualitative nature of the influence of visual attention on early visual processing. It remains elusive how load modulations of early visual cortices influence the content of processing of these brain areas. Investigating the representation of a basic visual feature in early visual cortices under different load modulations might offer an interesting chance to assess the exact nature of the influence of attentional load and thereby close this gap.

## I.2 Studying feature-specific processes in the human brain

### I.2.1 The representation of basic visual features in the brain

It is well known from electrophysiology research that in the primary visual cortex (V1) of primates, neurons show preferences in their response to different visual inputs. One well studied example is the so called orientation preference of neurons in early visual cortex (Hubel and Wiesel, 1962, 1968). This phenomenon describes differences in the response of neurons depending on the orientation of a line presented as visual input, i.e. a neuron might respond strongly to a horizontal line, but display a smaller response to a vertical line (Figure I-3 top). Neurons with different orientation preferences are systematically mapped across the cortical surface, with regions containing neurons of similar orientation preferences separated by approximately 500 µm (Wang et al., 1996, shown in Figure I-3 a, different colours represent different orientations). Thus, basic visual features like orientation are represented at a much finer spatial scale in the cortex than the resolution of a high-field fMRI image acquired in voxels. A voxel is the cube shaped, smallest region that an fMRI scanner is able to record activity from. The difference between a common voxel size of 3x3x3mm (indicated by the black grid in Figure I-3 a) and cortical regions showing an orientation preference (indicated by different colours in Figure I-3 a) is relatively large. It is thus remarkable that fMRI is



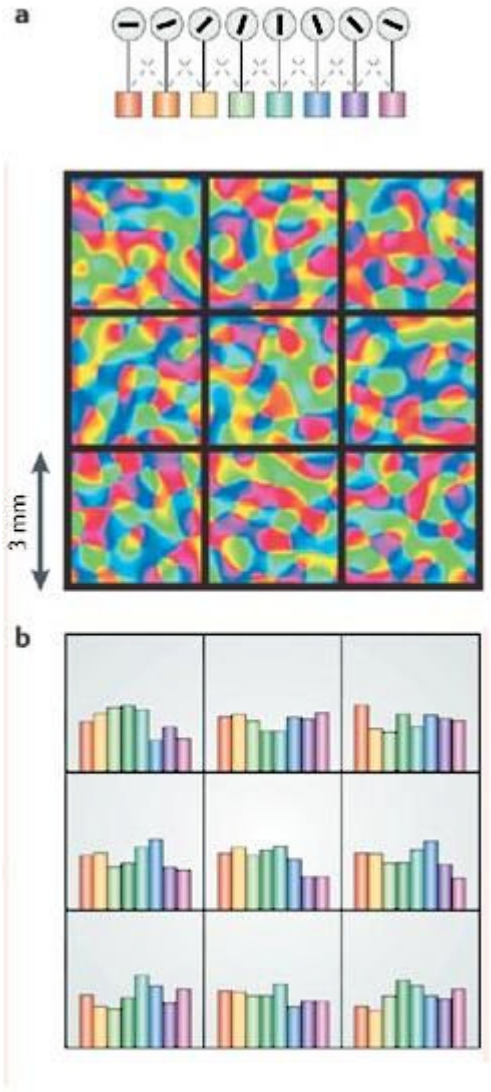




**Figure I-2**
**Top: Electrophysiology recordings from V1 like in Hubel and Wiesel (1962). One specific cell is responding strongest for only one orientation. The tuning of a single cell can be expressed with a tuning curve of that one cell (right).**
**Right: a) Responses to different oriented lines are recorded from many locations across primary visual cortex (V1). Orientation preferences of neurons are expressed with different colours. Neurons with different orientation preferences are systematically mapped across the cortical surface. Similar orientation tuning neurons group in regions of 500µm width. High field fMRI resolution of 3x3x3mm is illustrated with the black grid. b) Irregularities in the map of similar orientation tuning result in regions that over-represent one particular orientation. These potential biases in the orientation preference of each voxel can be simulated with a histogram of orientation tuning per voxel.**

none the less able to measure differences in underlying neural tissue represented at far smaller size.

### I.2.2 The emergence of multivariate analysis for fMRI data

In humans, Blood-oxygen-level dependent (BOLD) fMRI signals are typically measured at spatial resolutions far greater than the size of

single cells or even large groups of neurons in brain tissue (i.e. cortical columns), thus measuring neural representations at or below the level of large groups of neurons in brain tissue was traditionally thought inaccessible with fMRI. This belief was further corroborated by the fact that most fMRI analyses apply pre-processing steps such as spatial smoothing and spatial transformations that actively blur the responses recoded at a single voxel in order to get a better local area estimate. As a consequence, conventional analysis approaches generally focused on relatively large areas, often additionally averaging all activity recoded in one location over all times a particular condition has been recorded, in order to get a more reliable overall average.

It has recently emerged that fMRI can be used to study fine-grained neural representations, even when they are encoded at a finer scale than the resolutions of the measurement technique. By taking into account subtle biases in the pattern of activity recorded, measured simultaneously at many locations and at the very limit of spatial and temporal resolution, subtle biases in the pattern of activity have been demonstrated to allow the study of processes of the human brain thought to be represented below the spatial resolution of fMRI (Haxby et al., 2001; Cox and Savoy, 2003; Mitchell et al., 2003; Haynes and Rees, 2005b, 2005a; Kamitani and Tong, 2005b, 2005a; Haynes and Rees, 2006; Kamitani and Tong, 2006; Kriegeskorte et al., 2006; Serences and Yantis, 2006; Haynes et al., 2007; Serences and Boynton, 2007a; Formisano et al., 2008; Mitchell et al., 2008; Sumner et al., 2008). As a consequence, by taking into account the full spatial pattern of brain activity, so called

'multivariate' analysis, it is now generally deemed promising to study neural representations that have been previously thought inaccessible to non-invasive imaging techniques in humans.

While multivariate analysis has proven itself as very powerful tool, it can by no means entirely replace traditional univariate analysis techniques. On the contrary, both techniques address the same question by studying fMRI data differently: What can we infer about the underlying neural representation from an fMRI-signal? Hence, univariate and multivariate analysis are both complementary and independent analysis-tools and should both be viewed as fMRI analysis techniques each in its own right.

Comparing conventional 'univariate' analysis with 'multivariate' analysis reveals several differences. Conventional analysis often compares whether the average signal recorded from a voxel or region of interest (ROI) during one condition is significantly different from the average signal during a second condition, often by acquiring many samples of brain activity to maximize statistical sensitivity. However, multivariate analysis accumulates, from many voxels (spatial locations), the weak information available at each single location, instead of focusing on averaged activity. Additionally, contrasting to most conventional analyses, multivariate analysis only rarely employs pre-processing steps (such as spatial smoothing or normalization). This means that, conversely to univariate analysis, fine-grained spatial information that might carry information about perceptual states of an individual is not lost. In its most extreme form that means that multivariate analysis is able

to decode, quasi-online, estimates of a person's perceptual or cognitive state (Haynes and Rees, 2005b).

### I.2.3 Studying orientation and motion specific processing in early visual cortices with multivariate analysis

First examples examining basic visual feature processing in early visual cortices in humans include studies investigating the neural representation of motion directions and oriented lines (Haynes and Rees, 2005a; Kamitani and Tong, 2005b, 2005a). In these studies the authors show that, by using multivariate techniques, different instances of basic visual features such as different directions of motion or differently oriented lines result in different spatial patterns of activity in early visual cortices. Each condition (i.e. oriented lines tilted 45 degrees to the right versus oriented lines tilted 45 degrees to the left, Haynes and Rees, 2005a) can be distinguished from the other significantly above chance as a result of multivariate analysis. Such results have thus been termed feature specific 'decoding' of brain activity. Previous to these studies, research on visual feature specific processing in animals resulted in equally impressive results. Preceded the findings in humans, examples include imaging of monkey orientation columns (Wang et al., 1998) and highly detailed imaging of cat orientation columns (Kim et al., 2000; Kim and Fukuda, 2008).

In humans, further work enabled Kamitani and Tong to demonstrate feature specific neural patterns not only according to one of two

stimuli, but also to find specific patterns according to the attentional selection of one of two overlapping stimuli: in particular two orientations (Kamitani and Tong, 2005b) and two overlapping direction of motions (Kamitani and Tong, 2006). More recently Serences and Boynton (2007) also claimed to have measured the influence of feature-based attention on unattended stimuli and even unstimulated areas of visual cortex (Serences and Boynton, 2007a).

In summary, the use of multivariate analysis to decode different features or basic visual stimuli such as different orientations or different directions of motion is a very young sub-discipline of vision research using fMRI. However, it has the potential to widen our understanding of basic visual processing in the human brain substantially and, moreover, offers a variety of new directions.

### I.2.4 New directions in the study of feature specific processing in human visual cortex

The emergence of multivariate analysis techniques for fMRI has led to a number of groundbreaking in-vivo demonstrations of visual features represented in the human visual cortex. It is interesting to note that many new findings are facilitated not by using multivariate analysis alone, but often in conjunction with conventional, univariate analysis. Beyond the initial findings on orientation and direction of motion (see above, Haynes and Rees, 2005b, 2005a; Kamitani and Tong, 2005b, 2005a), also other categories of stimuli were successfully analyzed with multivariate

analysis, including spatial frequency (Bahrami, 2009, submitted), joint sensitivity to both color and orientation (Sumner et al., 2008), color alone and in conjunction with motion (Seymour et al., 2009), different object categories (O'Toole et al., 2005) and even black and white natural images (Kay et al., 2008). Thus, for all of the above stimulus categories it holds true that they are represented in an anisotropic way in the human cortex. In some cases speculations of an underlying columnar representation are likely (i.e. for spatial frequency or color). Following from these and other technical demonstrations of feature-based specific decoding, new directions arise: On the one hand side these new opportunities include further visual stimuli that have so far not been shown to be represented in a way fMRI analysis can access. Examples could include the representation of different categories of surfaces or basic shapes but also slightly more complicated visual stimuli like different types of places (i.e. outdoor vs. indoor), faces (i.e. male vs. female) or specific object categories.

Yet, it might be even more fruitful to utilize multivariate analysis as a dependent measure to differentiate between different types of brain states. By combing multivariate analysis techniques with basic psychological questions like afore mentioned load theory, for example, interesting new insights might be gained. As an illustration, in such a paradigm multivariate measures might give an alternate insight into the underlying brain activity of a distractor under differing load conditions, additionally to a univariate main effect or interaction. Thus, multivariate analysis might offer further

insight into how basic visual features are represented under different attentional conditions.

Another new direction in the application of multivariate techniques could be the study of the representation of unconsciously processed stimuli. In this respect, Haynes and Rees (2005) demonstrated that specific orientation information could be read out from brain activity with multivariate analysis even though it was rendered invisible to participants (Haynes and Rees, 2005a). In another study multivariate analysis revealed that unconsciously perceived (invisible) faces and houses are still represented in the neural activity of the fusiform face area and the parahippocampal place area (Sterzer et al., 2008).

Overall, especially the combination of univariate and multivariate analysis techniques provides a potentially powerful toolset to explore new and exciting questions ever getting closer to understanding the nature of the underlying neural representation of the BOLD signal.

## I.3 Feature-based Attention

### I.3.1 Feature-based attention versus spatial attention

Attention modulates the level of brain activity and behavioral performance associated with the processing of attended visual stimuli. When multiple stimuli are simultaneously present in a scene, they compete for cortical representation and access to awareness (Desimone and Duncan, 1995; Serences and Yantis, 2006). Thus, selecting which information to attend to in a visual scene is a crucial aspect of sensory processing. It is widely believed that, based on current behavioral goals, relevant stimuli are processed more efficiently than irrelevant stimuli. To achieve an advantage to stimuli presented at the selected location, an observer can attend to a particular region of space, commonly referred to as spatial attention (Moran and Desimone, 1985). Yet, in everyday life, humans often know more about the defining features of objects (e.g. "the pink post-it note") than precise spatial locations ("where the post-it note is on your refrigerator-door"). In order to achieve an advantage to stimuli with known features rather than known location, spatial attention is of little help. Conversely, humans will attend to a particular visual feature (in the example above the color pink). This type of selective attention is commonly referred to as feature-based attention (Treue and Maunsell, 1996; Treue and Martinez Trujillo, 1999; Martinez-Trujillo and Treue, 2004).

### I.3.2 Effects of Feature-based attention in monkeys

*Area V5/MT and its relevance to feature-based attention research*. Feature-based attention on brain activity has, until recently, been mostly addressed in animal research, typically using electrophysiological techniques in monkeys. Single-unit studies in monkeys have shown that attention modulates neuronal signals in a range of areas in visual cortex (Maunsell and Treue, 2006). A majority of monkey research on feature-based attention has relied on analyzing neuronal responses from one specific, motion sensitive, area in the macaque monkey brain (Allman and Kaas, 1971, Baker et al., 1981). Zeki et al. (1983) later defined an area selectively responsive to motion in the macaque brain (Zeki, 1983; Zeki et al., 1991). Zeki labeled this area "V5". However, around the same time a similar result was published by another group, and instead of calling the area "V5", this motion responsive area was labeled "MT" for middle-temporal area, due to its cortical arrangement (Albright et al., 1984). Further research established both areas as homologous and, thus, for the remainder of this thesis it is referred to as V5/MT. Soon after the initial discovery of monkey area V5/MT, it was found to represent different directions of motion in a columnar architecture (Tootell et al., 1995; Tootell and Taylor, 1995).

Measuring responses from neurons in motion sensitive area V5/MT, Treue and Martinez-Trujillo demonstrated that neurons tuned to the attended feature show an enhancement (gain) of responses and a suppression of the opposite feature (Treue and Martinez Trujillo,

1999). This increase and decrease of neuronal activity has been termed the 'feature-similarity-gain' of the response of a particular neuron or the population response of many neurons. Gain changes are described as an enhancement of the selectivity of the population response that emphasizes the attended over an unattended (Martinez-Trujillo and Treue, 2004). The feature-similarity gain mechanism modulates the firing rate of neurons tuned to an attended feature when a neurons receptive field is inside the current location of spatial attention (Treue and Martinez Trujillo, 1999) and also when neurons are driven by a stimulus outside the focus of spatial attention (Martinez-Trujillo and Treue, 2004). As a result, the two prominent findings from the study of feature-based attention are the modulation of neurons whose receptive field overlaps the current spatial focus of attention and the modulation of neurons that have their receptive field outside the current focus of spatial attention (Treue and Martinez Trujillo, 1999; Martinez-Trujillo and Treue, 2004; Bichot et al., 2005; Womelsdorf et al., 2006). In conclusion, it is widely accepted that feature based attention is independent of spatial attention: It acts directly on tuned neurons irrelevant of their retinotopic representation of visual space (the spatial location they represent) and importantly even if spatial attention is directed elsewhere.

### I.3.3 Feature based attention in humans

Soon after the discovery of a motion sensitive area V5/MT in monkeys, a counterpart in the human brain was revealed, mainly

from fMRI-adaptation studies (Heeger et al., 1999; Nishida et al., 2003; Seiffert et al., 2003). Today, area V5/MT is well established in the human and monkey brain as a motion responsive area with two distinct subregions: MT and MST (Dukelow et al., 2001; Huk et al., 2002). While area MT responds mostly to contralateral motion-stimulation, MST responds mainly to ipsilateral motion stimuli. Due to the lack of single cell recoding in humans and the coarse spatial resolution of non-invasive imaging techniques such as fMRI, the study of feature-based attention in humans has traditionally been thought inaccessible (for a rare exception see Saenz et al., 2002). However, the emergence of multivariate analysis rapidly changed this view and the neural representation of basic visual features in humans is now an active area of research (for an early review see Haynes and Rees, 2006). Moving from the neural representation of basic visual features to feature-based attention enabled Kamitani and Tong to find feature specific neural patterns according to the attentional selection of one of two overlapping orientations (Kamitani and Tong, 2005b) and direction of motion (Kamitani and Tong, 2006). Another study measured a behavioral response (the tilt aftereffect resulting from adaptation to a set of oriented lines) and feature specific brain activity (of one orientation) and found reduced activity for an unattended feature as opposed to the attended feature (Liu et al., 2007). However, importantly, none of the afore mentioned studies specifically demonstrated the influence of feature-based attention to spatially unattended areas in humans - a key aspect for establishing the independence of spatial and feature-based attention (see Kanai et al., 2006, for another

suggestion towards the independence of spatial and feature based attention).

In 2007, Serences and Boynton demonstrated the independence of feature-based attention and spatial attention in humans (also see Saenz et al., 2002; Serences and Boynton, 2007a): They looked for specific spatial patterns of fMRI signals as a function of feature-based attentional selection of one of two overlapping motion directions. The independence of feature-based and spatial attention was shown by a spread of the influence of an attended feature from the neural representation of the attended stimuli to neural representations of unattended stimuli somewhere else in the visual field. Additionally, their experimental design also allowed restriction of stimulus evoked activity to one hemisphere only by showing their stimulus only on one side of the screen. In this setup, the underlying neural representation of the unstimulated hemifield showed a measurable modulation according the (feature-based) attentional selection in the attended visual hemifield (Serences and Boynton, 2007a). Thus, the result by Serences and Boynton (2007) is a qualitative replication of earlier findings from monkey single cell recordings (Treue and Martinez Trujillo, 1999). However, a closer observation of the distinction of feature-based and spatial attention as demonstrated by by Serences and Boynton (2007) reveals some important shortcomings (Kaul and Bahrami, 2008). These shortcomings are mostly due to possible alternate explanations of the results (see Chapter VI for details).

In conclusion, the study of feature-based attention in humans showed promising results pointing towards largely similar mechanisms of feature-based attention in humans and monkeys.

However, some details are yet to be proven without the possibility of alternate explanations.

### I.3.4 New directions of the study of feature-based attention in humans

The combination of a neural property well studied and understood in electrophysiological terms in monkeys and now accessible with non-invasive techniques in humans makes the study of feature-based attention an interesting candidate to further understand the nature of the underlying neural processes of fMRI recordings. So far, for attended stimulus representations, it seems that the neural mechanisms of feature-based attention in humans work similar to those found in monkeys (Kamitani and Tong, 2006; Liu et al., 2007). However, new studies could explore the influence of feature-based attention independent from spatial attention without the caveats of the aforementioned study by Serences and Boynton (2007).

Other promising routes in the study of feature-based attention might be to investigate the interplay of (feature specific) working memory and feature-based attention. For example, it has recently emerged that early visual areas can retain specific information about attended visual features held in working memory, even when no physical stimulus was present (Harrison and Tong, 2009; Serences et al., 2009a). Another study established a feature-specific effect independently for feature-specific working memory

enhancement and feature-based attention enhancement. Then, the authors investigated whether, when combined, the effects of feature-based attention and feature specific working memory occurred independently of each other or interact. While the behavioral findings show that both effects can independently modulate motion perception in humans (Mendoza et al., 2009), evidence for what happens with the underlying brain activity during such independent modulations remains elusive.

A final possible new direction for the study of feature-based attention in humans could present itself in the testing of theoretical models of attention. One interesting example is the recently published normalization theory of attention (Reynolds and Heeger, 2009). The normalization theory is especially interesting as it unifies many seemingly conflicting attentional models of the past including studies of spatial attention and feature-based attention studies and the concept of the feature-similarity-gain. So far, it has only been tested in the spatial attention domain (Herrmann et al., 2009), but if it could be corroborated in the feature-based attention domain too, this would significantly strengthen the model and our understanding of attention in general.

## I.4 Face processing in the human brain

### I.4.1 Studying higher order stimuli in the human brain

Visual perception has been studied at multiple levels of neuroanatomy and with a wide variety of stimuli. In general, the more complex visual stimuli become, the higher up in the anatomical hierarchy of brain areas we have to look to find brain activity specific for that type of stimulus (for review see Grill-Spector and Malach, 2004). For example, afore mentioned were basic visual features such as orientation, spatial frequency, color or direction of direction of motion which have been directly or indirectly shown to modify brain activity in very early steps of visual processing: spatial frequency, orientation and color as early as the LGN and V1 or direction of motion a little later, usually in V3a and V5/MT but also as early as V1 and V2 (most recent basic visual feature specific fMRI results include for example Singh et al., 2000; Kamitani and Tong, 2005b; Bartels et al., 2008; Sumner et al., 2008; Bahrami et al., 2009; Seymour et al., 2009).

Yet, especially humans are adept and specialized in far more than the perception of basic visual features. In everyday life, humans are surrounded by objects, other humans, animals, etc in complex surroundings like indoors, in motion, underground, etc. As humans we have developed an extremely versatile vision apparatus allowing us to accurately perceive all these complex settings. The wealth of studies of neuronal correlates of the perception of higher order stimuli is immense and has resulted in a number of remarkable

findings (for review see Grill-Spector and Malach, 2004). Most prominent stimuli-selective findings include object-selective areas (Malach et al., 1995; Grill-Spector et al., 1998), face selective areas (Puce et al., 1995; Kanwisher et al., 1997; Ishai et al., 2000; Grill-Spector et al., 2004) and place selective areas (Aguirre et al., 1998; Epstein and Kanwisher, 1998; Ishai et al., 1999) amongst others. While all these categories must be studied separately, there may be general organizational principles: for objects, for example, studies indicate that object-selective cortex encode objects according to a hierarchy, with stimulus-based representations in posterior regions and subjective representations in anterior regions (Haushofer et al., 2008). In this thesis, however, object-specific studies focus on face perception.

### I.4.2 Neuronal correlates of face perception in humans and monkeys

Humans are experts in face perception. With seemingly endless capacity humans can distinguish individual faces with high precision (Diamond and Carey, 1986). This made face perception a naturally interesting topic for neuroscience and indeed, much about the neural correlates of face perception is well studied: In both humans and monkeys, fMRI studies reveal a network of cortical regions that show increased blood flow when participants view images of faces, compared with other stimuli (for a recent review see Tsao and Livingstone, 2008).

Face processing is thought to be distinct from non-face object processing because it is said to be more 'holistic': faces are represented as non-decomposed wholes, rather than as a combination of independently-represented component parts (eyes, nose, mouth), and the relations between them (Farah et al., 1998). Faces are unique in the high degree to which they are processed holistically (but see Gauthier and Tarr, 2002 for other categories of holistic stimuli). A straightforward assumption about the neural correlates of face perception was that there must be a specific mechanism or area facilitating this unique, holistic processing. Puce et al. (1995) was first to demonstrate such a face-selective area, the fusiform gyrus, FG, often also labeled fusiform face area, FFA, with fMRI in humans, corroborated shortly after by other highly convincing results (Puce et al., 1995; Kanwisher et al., 1997). While fMRI signals are an indirect measure of brain activity, direct evidence towards the involvement of another occipital-parietal area comes from repetitive transcranial magnetic stimulation (TMS) targeted at the right inferior occipital gyrus, IOG, also labeled occipital face area, OFA (Pitcher et al., 2007). In humans, the FG and the IOG together are sometimes referred to as first stages in face-processing models (Haxby et al., 2000; Calder and Young, 2005). Recent results from single-cell recordings in monkeys substantiate the view of at least one highly face-specific area in monkeys: Tsao and colleagues demonstrated direct evidence towards a ventral temporal face area nearly exclusively selective to face-processing in the brain of monkeys (Tsao et al., 2003; Tsao et al., 2006). They found this area containing visually responsive neurons which were strongly face selective (97% of neurons)

indicating that a dedicated cortical area exists to selectively support (holistic) face processing. While the existence of this area does not undermine the proposed existence of a cortical network of face processing areas, its exclusiveness poses a challenge to a more modular view of face and object processing in humans. For example, Haxby et al. argued that objects and faces are coded in a large network of areas and also via the distributed profile of neuronal activity across the network of areas with much of the ventral visual pathway involved and not confined to a single area alone (Haxby et al., 2000; Haxby et al., 2001). Over the years many human fMRI studies of face perception concluded that faces are processed in a distributed network of brain areas rather that in one or two single, specialized areas (Haxby et al., 2000; Ishai et al., 2005; Fox et al., 2008; Ishai, 2008). The emergence of multivariate techniques in fMRI research might help answer some of these seemingly contradictory results by providing a new approach to study specific features of face processing.

### I.4.3 The core and extended network of face processing

Many human imaging studies of face perception converge on the conclusion that faces are processed in a distributed network of brain areas (Haxby et al., 2000; Ishai et al., 2005; Fox et al., 2008; Ishai, 2008). A "core system" has been proposed, comprising of three regions that mediate the analysis of invariant facial features: the (FG), the inferior occipital gyrus (IOG) and the posterior superior temporal sulcus (STS) (Ishai et al., 2005; Gobbini and

Haxby, 2007). Additionally, the "extended system" includes regions that mediate the processing of changeable aspects of faces, such as mood and expression. The extended system includes limbic regions, such as the amygdala (AMG) and insula (Ishai et al., 2004; Ishai et al., 2005); the inferior frontal gyrus (IFG) (Ishai et al., 2005), and regions of the reward circuitry, especially the nucleus accumbens and medial orbitofrontal cortex (OFC) (Aharon et al., 2001; Ishai, 2007).

Many regions of the core and extended systems display greater brain activity when specific aspects of face processing are required by task demands. For example, the FG/FFA and IOG are more active in processes that require the identification of individuals (Puce et al., 1995; Kanwisher et al., 1997; Ishai et al., 2000; Grill-Spector et al., 2004). Gaze direction and speech related movements seem to be processed in STS (Puce et al., 1998; Calder and Nummenmaa, 2007). The amygdala and insula are implicated in processing faces with emotional context and facial expressions (Breiter et al., 1996; Vuilleumier et al., 2001; Ishai et al., 2004; Ishai et al., 2005), the IFG is activated during the processing of semantic aspects (Ishai et al., 2000; Leveroni et al., 2000) and finally the OFC is processing facial beauty, sexual relevance and reward value (Aharon et al., 2001; O'Doherty et al., 2003; Kranz and Ishai, 2006; Ishai, 2007).

### I.4.4 Specific brain responses to features of faces

More recently, another fMRI study further illustrated the functional division of labor between different parts of the proposed core face network (FG, IOG and STS) and offered an interesting solution for the perceived discrepancy between single-unit recordings and fMRI measurements. Using stimuli consisting of face parts and the configuration of those parts the IOG and the STS was sensitive to the presence of face parts but not their spatial configuration as a face (Liu et al., 2009). However, the FG was sensitive to both kinds of information and only in the FG was the response to configuration and part information correlated across voxels. Thus the FG may contain a unified (holistic) representation of faces including both kinds of information. However, while the idea that the IOG conducted an earlier stage of face processing than the FG was consistent with its location posterior to the FG, more importantly, it implied connectivity between the different face-selective regions in the human ventral visual pathway. It thus supported the notion of a distributed face network for different (sub-) functions of face processing and, additionally, a highly specialized area for face processing.

One obvious function of face processing in humans must be the discrimination of facial gender, a very common task in everyday face perception. However, although gender is a fundamental characteristic of faces, conventional fMRI data analyses have not localized a region within the face network specialized for the discrimination of gender. Investigating fMRI adaptation to facial

gender and race, one study showed the strongest adaptation effects outside the face network, in the cingulate gyrus, while the same subjects showed only weak adaptation effect in regions of the core face network (Ng et al., 2006). Another study, looking specifically at gender-related face processing, found differences in brain activity related to sexual preference of the participants (Kranz and Ishai, 2006). In this fMRI study, forty hetero- and homosexual men and women viewed or rated the attractiveness of male and female faces, in order to test whether they would respond more to their sexually-preferred faces. A significant interaction was found between stimulus gender and the sexual preference of participants in the thalamus and medial orbitofrontal cortex: Heterosexual men and homosexual women responded more to female faces, whereas heterosexual women and homosexual men responded more to male faces. However, importantly, despite the large number of subjects and corresponding statistical power, no face-gender difference was found in early regions of the (core) face network. Thus, it remains elusive whether or where the representation of facial gender is localized in the human brain.

### I.4.5 New directions in feature specific face perception research

Conventional univariate analysis leaves an incomplete picture about the representations of specific features of face processing in the human brain. However, feature specific brain activations have been successfully distinguished with multivariate pattern analysis. Thus,

the study of such feature specific representations of faces is a potential new direction for fMRI studies of face processing; especially face perception as such has already been shown to result in a distributed spatial pattern of activity (Haxby et al., 2001).

Instead of searching for a specific function of face processing, one very recent study rather looked at distinctive fMRI patterns during the performance of different face-related tasks (Chiu et al., 2009). Applying multivariate analysis showed that regions of the core network (FG and IOG) showed task-specific modulation during race vs. gender categorization. In this study, a differential spatial pattern of activity was found depending on whether participants performed a race discrimination task (caucation vs. asian) or a gender discrimination task (female vs. male). While stimuli remained constant, the two task instructions alone were sufficient to modulate the spatial activity pattern in IOG and FG such that the two tasks could be told apart from each other significantly more often than chance performance. Yet, the setup of different tasks was not ideal to infer specific functions of face processing; still this study illustrated that different attentional conditions sufficed to modulate activity in these areas feature specifically – thus these features might be represented somehow in these areas. Further research might simply measure the brain pattern response to different races or gender and compare these to fully answer this question.

## I.5 Summary and outlook

In this thesis, I investigate basic questions about visual processing, with one central question in mind: What are the current limits of our knowledge of brain activity underlying vision and can I further this knowledge? I selectively reviewed the research that prompted specific experimental questions addressed in this thesis, starting with the longstanding debate regarding the selection process of visual attention (early or late) and its potential solution by load theory. But load theory does not answer questions about the specifics of how distractor brain activity is modulated by different attentional load conditions. However, the study of feature specific brain activity (with the help of multivariate analysis) might offer a chance to explore one such specific aspect of brain activity. In chapter III, IV and V, I combine load theory with multivariate analysis to explore the specific question whether attentional load alters the representation of basic visual features in early visual cortices.

Following load theory and the study of basic visual features in humans, I outline the evidence of the independence and importance of feature-based attention. Accordingly, chapter VI presents an experimental paradigm to test my hypothesis that the influence of feature based attention in early visual cortices is far more pronounced than previously thought.

Finally, I describe face specific processing in the human brain and evidence towards both the specific involvement of single areas and

a wide-spread face responsive network. In view of this, chapter VII presents a study about the distribution facial gender information in the human brain.

However, before the experimental sections, chapter II first reviews, in detail, methods and techniques used throughout the experimental work described in the subsequently presented studies.

# Chapter II:

# General Methods

This chapter describes MRI techniques that form the methodological basis to all chapters presented in this thesis. I give a brief, non-technical, overview about the general background and function MRI and functional MRI (fMRI). I follow with a detailed description of fMRI data-analysis with statistical parametric mapping (SPM) plus functional area localization with retinotopic mapping. Multivariate analysis is covered in the final section of this chapter.

## II.1 Functional MRI

### II.1.1 A brief overview of the physics of MRI

When placed in a uniform magnetic field (B0 field), spins of unpaired atomic nuclei (mainly protons) contained in any object (for example a participant's brain) align parallel to the B0. Additionally, the B0 field, in combination with proton angular momemtum, forces the spinning protons to revolve, or precess, around the axis of this magnetic field at a frequency proportional to the strength of the B0 (known as the resonance frequency). The direction of the precessing around the main field direction of the B0 is random for all nuclei resulting the overall transverse magnetization (TM) to be zero. Applying a perpendicular radio frequency pulse to the B0 at the resonance frequency causes nuclei to absorb this energy and their spins to move away from their equilibrium positions. As a

direct result, the current local magnetization is not perfectly aligned anymore (to the B0), but tilted towards the newly applied radio frequency magnetic field. After the radio frequency pulse, protons relax again and realign with the UFM - this realignment process emits energy. In an MRI scanner this emitted energy is recorded in a receiver coil surrounding the object scanned (Squire and Novelline, 1997).

### II.1.2 Formation of images using MRI

To create an image with MRI, tissue differences have to be distinguishable on the basis of their spatial location, in other words their X, Y and Z components in space. In most modern MRI scanners this problem is solved by a multitude of methods. Assuming the B0 alignment forms a Z-axis, the TM will be tilted in the XY plane (towards the radio frequency magnetic field). Resolution along the z-axis is created by exciting the sample only one slice at a time, by combining the frequency gradient with a radio frequency pulse of a particular frequency and bandwidth. Heavily depending on the type of image acquired, common slice thicknesses lie between 0.75-5mm with varying distances between these slices (usually zero to 1mm). Within each slice the X and Y components are mainly determined with the help of gradient coils. Gradient coils are resistive electromagnets powered by sophisticated amplifiers. They permit rapid and precise modifications to their magnetic field strength and direction. Combined with the large B0, gradients coils produce a systematic modification (a linear gradient) in the magnetic field. Orthogonal

gradients can be combined freely and form the basis for additional components of the coordinate system. The magnitude of the gradient allows encoding of position along one axis (the x-axis). Phase encoding enables encoding of position in a second dimension (along the y-axis). As a result, discrete increases in the frequency encoding and phase encoding gradients divide each slice into small cubes, called voxels (volume elements). All the protons in a voxel experience the same frequency and phase encoding, and the signal from a voxel is the sum of the signal for all the protons in that voxel.

Contrast in each image voxel is created by the differences in signal intensity from different tissues. The largest contribution to the signal comes from hydrogen atoms in the water components of tissue (or other biological fluids: i.e. blood has a high hydrogen concentration). Thus, signal intensity depends in part on the density of hydrogen atoms. Different properties of the relaxation times of different tissues combined with varying echo and repetition times of the radio frequency pulses result in a multitude of possibilities for image acquisition. Most commonly used are so called T1-weighted MRI sequences which provide good contrast of white and gray matter in the brain. T1-images are the most commonly used clinical scans. For functional MRI image series, however, T2*-weighted MRI images are generally acquired. T2* images provide the highest contrast for different levels of blood oxygenation levels. Echo-planar imaging (EPI), allows extremely rapid acquisition of whole T2*-weighted brain images. While many MRI sequences can only acquire partial data per slice per RF pulse, EPI sequences acquire an entire

slice-data-set after each RF pulse. As a result, an image of a complete slice (along the z-axis) can often be acquired in less than 100ms. This means that acquisition time is far lower for EPI, making it very suitable for recording dynamic information, like in functional MRI. Each single data point acquired during a 2D MRI sequence contains information about the entire slice. Voxels are reconstructed by summing information from many such data points, each of which can be thought of as points in spatial frequency space (k-space), and each of which contain information about the entire slice. All the fMRI experiments in this thesis use EPI sequences of T2*-weighted images and focus on blood-oxygen-level depended (BOLD) signal changes.

In modern fMRI experiments, the researcher predetermines the number of slices along with the slice thickness, inter-slice distance and the in-plane resolution. The number of slices times the time it takes to acquire each slice results in a total acquisition time per volume - also called TR. The TR is generally seen as the temporal resolution of an fMRI experiment as any voxel is covered exactly once per volume, or once every TR.

### II.1.3 The BOLD signal

In general, neuronal activity and increased local glucose metabolism are tightly coupled to a local increase in blood flow. fMRI measures neural activity indirectly by detecting changes in regional blood flow as indicated by blood oxygenation levels. The MRI signal is sensitive

to the oxygenation state of haemoglobin as deoxyhaemoglobin is more paramagnetic than oxyhaemoglobin (Pauling and Coryell, 1936). Paramagnetic substances have a more rapid transverse relaxation time, and a shorter T2* time constant, resulting in a reduced T2* weighted MRI signal. Thus deoxyhaemoglobin produces a smaller MRI signal than oxyhaemoglobin. It is generally accepted that this is what principally underlies the BOLD signal (Logothetis et al., 2001). It was first discovered in mice that blood with more deoxyhaemoglobin will produce a reduced signal relative to highly oxygenated blood (Ogawa et al., 1990); subsequently in human visual cortex (Kwong et al., 1992). BOLD contrast is determined by the balance between supply, determined by blood flow and blood volume, and demand, determined by the surrounding tissue's rate of glucose metabolism, and consumption of oxygen. Local increases in neural activity lead to increased glucose metabolism and increased oxygen consumption (Vanzetta and Grinvald, 1999). As the rise in oxygen uptake is smaller than the rise in blood flow to activated brain regions (Fox and Raichle, 1986), there is an overall increase in blood oxygenation levels lasting for several seconds. This overcompensation is the basis for the increased BOLD signal seen when neural activity increases. This increase in BOLD contrast, caused by the decrease in deoxyhaemoglobin and measured in fMRI, is delayed in time with respect to the neural activity. Typically the BOLD signal peaks 4-8 seconds after the onset of neural activity. The rise and subsequent return to baseline of the BOLD signal is known as the haemodynamic response function (HRF).

### II.1.4 Neural correlates of the BOLD signal

The precise relationship between the underlying neural activity and BOLD is under active research (e.g. Logothetis et al., 2001; Logothetis and Wandell, 2004). Especially, the specific cellular and molecular mechanisms underlying the BOLD signal are still a matter of debate. Most importantly, it remains unclear whether the BOLD signal is mainly correlated to spiking activity of neurons (output) or synaptic activity (input).

From a physiological point of view, it is widely believed that increased blood flow follows directly from increased synaptic activity, as blood flow increases in proportion to glucose consumption (Fox and Raichle, 1986) and glucose metabolism is linked to synaptic activity (Rothman, 1998). Thus, the hemodynamic response might primarily reflect the neuronal input to the relevant area of the brain and its processing there rather than the long-range signals transmitted by action potentials to other regions of the brain (Logothetis and Wandell, 2004). As a direct result this would mean that in situations when input into a particular area plays a primarily modulatory role, fMRI experiments may measure activation that does not correlate well with single-unit measurements.

However, in order to truly quantify the neural basis of the BOLD signal, one promising possibility is the direct comparison of electrophysiological studies and fMRI studies. For example, a comparison of single unit data from monkey V5/MT (a motion

responsive cortical area) with human fMRI measurements from V5/MT (the human homologue) showed that neuronal firing and BOLD responses increased linearly with increasing motion coherence (Rees et al., 2000). This is consistent with the two measures being well correlated. Furthermore, simultaneous recording of multi-unit activity (MUA) and local field potential (LFPs) from microelectrodes placed in monkey primary visual cortex while measuring BOLD contrast responses using fMRI (Logothetis et al., 2001) has broadly shown good correlation between these measures, indicating a high correlation with output spiking activity. However, for the whole brain this correlation was variable and highly dependent on the brain area considered. On average, LFPs correlated slightly more than MUA with the BOLD response. MUA represents the spiking activity of neurons near (~200μm) the electrode tip, while LFPs reflect synchronized dendritic currents averaged over a larger volume of tissue (reflecting inputs and intracortical activity), and they often (but not always) correlated with output spiking activity.
In conclusion this suggests that the BOLD response probably reflects components of both spiking output and synaptic input activity. Further research will be necessary to determine this issue in more detail.

### II.1.5 Limitations of BOLD dependent fMRI

Limitations on the spatial and temporal resolution of fMRI are of a physiological nature, imposed by the spatio-temporal properties of the HRF amongst other factors (Friston et al., 1998b; Logothetis,

2008). The BOLD signal originates in red blood cells in capillaries and veins surrounding the activated neural tissue, and thus is an indirect measure of tissue oxygenation and neural activation; thus the maximum spatial resolution obtainable with the BOLD signal is dependent on the local structure and density of the vasculature in a particular brain region. Due to local differences in these factors, the overall magnitude of the fMRI signal could potentially be misleading when comparing differences between brain regions. Additionally, the fMRI signal may not distinguish excitation from inhibition. Still, in a recent review Nikos Logothetis concludes: "fMRI is currently the best tool we have for gaining insights into brain function and formulating interesting and eventually testable hypotheses, even though the plausibility of these hypotheses critically depends on used magnetic resonance technology, experimental protocol, statistical analysis and insightful modeling" (Logothetis, 2008).

## II.2 fMRI Analysis - Preprocessing

In all experiments presented in this thesis, Statistical Parametric Mapping software (SPM, www.fil.ion.ucl.ac.uk/spm) was used to perform the analyses. SPM is a software package implemented in MatLab that allow preprocessing of fMRI data into a form that can then undergo statistical analysis with SPM to look at the effect of experimentally manipulated variables. fMRI data analysis with SPM can be dissected into discrete steps. Here, I selectively describe steps common to all of the experiments presented in this thesis. I

divide this description in two logical elements: preprocessing and statistical analysis.

### II.2.1 Spatial realignment

FMRI data is commonly acquired in a number of sessions (sometimes also called runs or scans). Experiments in this thesis usually have around 10 sessions per subject, each lasting around 3-5 minutes. Every session consists of a number of scan volumes. A common starting point for analysis of fMRI studies is to manually discard the first few image volumes from each session to allow for magnetic equilibration effects.

What follows is the first preprocessing step: spatial realignment. During a scan, head motion causes changes in signal intensity of a voxel over time, due to participants moving their head in space. Despite head restraints, most subjects will move their heads at least a few millimeters. Realignment involves applying an affine rigid-body transformation to align each scan with a reference scan (usually the first scan or the average of all scans) and resampling the data using tri-linear interpolation. The 6 parameters of the rigid-body transformation, representing adjustments to pitch, yaw, roll, and in X, Y, Z position, are estimated iteratively to minimize the sum of squares difference between each successive scan and the reference scan (Friston et al., 1995). However, even after realignment significant movement related signals persist (Friston et al., 1996). This is due to non-linear effects, including movements

between slice acquisitions, interpolation artifacts, non-linear distortion of magnetic field and spin excitation history effects, which cannot be corrected using an affine linear transformation. These non-linear movement related effects can be estimated and subtracted from the original data by including the estimated movement parameters from the realignment procedure in the design matrix during the model estimation stage (see below) of the analysis (Friston et al., 1996). At the end of spatial realignment, all data is commonly saved in the space of the reference scan, the resampled format.

### II.2.2 Spatial Coregistration

Following spatial realignment within sessions, a second crucial step is to ensure that all sessions relevant to a particular study are analyzed in the same anatomical space. For any subject-specific analysis techniques it is important that every session is co-registered with the same anatomical space. This is true for all functional scans, but also for structural scans or region-of-interest (ROI) mask images. In most studies of this thesis, the mean functional image of the main experiment (created during realignment) was used to establish a common space for all other images to get co-registered to. The warping parameters that map any other image onto this mean image space are modeled as a 12-parameter affine transformation, where the parameters constitute a spatial transformation matrix. The parameters are estimated iteratively, within a Bayesian framework, to maximize the posterior

probability of the parameters being correct. The posterior probability is the probability of getting the given data, assuming the current estimate of the transformation is true, times the probability of that estimate being true (Ashburner and Friston, 1997). Finding this solution involves jointly minimizing the sum-of-squares differences between the image to be coregistered and the mean functional image, and the prior potentials, which are used to incorporate prior information about the likelihood of a particular warp. In the case of multiple images in another space in need of coregistration (i.e. another function scan), the estimated warp can be applied to any number of the other images.

## II.3 fMRI Analysis – statistical parametric mapping

The approach used by SPM for analysis of fMRI data is based on the conjoint use of a General Linear Model (GLM) and Gaussian Random Field (GRF) theory to test hypotheses and make inferences about spatially extended data through the use of statistical parametric maps. The GLM is used to estimate parameters for the variables that could explain the BOLD signal time series recorded in each and every voxel individually. The resulting statistical parameters determined at each and every voxel are then assembled into three-dimensional images – Statistical Parametric Maps (SPM), that can then be contrasted with one another. Gaussian Random Field theory is used to resolve the problem of multiple comparisons that occurs when conducting statistical tests across the whole brain. The voxel

values of the SPM are considered to be distributed according to the probabilistic behavior of Gaussian fields, and 'unlikely' excursions of the SPM are interpreted as regionally specific effects, caused by the experimentally manipulated variables.

### II.3.1 General Linear Model

The general linear model is used in SPM to partition the variance in the observed neurophysiological response into components of interest, i.e. the experimentally manipulated variables, confounds and error, and to make inferences about the effects of interest in relation to the error variance. For each voxel the GLM explains variations in the BOLD signal time series (Y) in terms of a linear combination of explanatory variables (x) plus an error term (ε):

$$Yj = xj1\beta1 + xj2\beta2 + \ldots\ldots + xjl\betal + \ldots\ldots + xjL\betaL + \varepsilon$$

The β parameters reflect the independent contribution of each independent variable, x, to the value of the dependent variable, Y, .i.e. the amount of variance in Y that is accounted for by each x variable after all the other x variables have been accounted for. The errors, ε, are assumed to be identically and normally distributed. The GLM can also be expressed in matrix formulation:

$$Y = X\beta + \varepsilon$$

Where Y is a vector of J BOLD signal measurements (one per image volume) at a particular voxel (Y = [1…j…J]) and β is the vector of the parameters to be estimated (β=[ β1… βj… βJ]. X is the design matrix containing the variables which explain the observed data. The matrix has J rows, one per observation, and L columns, one per explanatory variable (x) (also referred to as covariates or regressor). The regressors, which form the columns of the design matrix (and have one value of x for each time point j), are created for each explanatory variable manipulated in the experiment (the experimental conditions) by placing delta functions at the time points corresponding to the events of interest and convolving this vector with the haemodynamic response function. The HRF is modeled in SPM with a multivariate Taylor expansion of a mixture of gamma functions (Friston et al., 1998b; Friston et al., 1998a).

Movement parameters, calculated during realignment, can be including in the model as additional regressors to account for movement artifacts which are not corrected by realignment itself. Temporal confounds must also be eliminated from the data. Prior to fitting the model a high pass filter is applied to the data to eliminate drifts in the magnetic field and the effects of movement. A low pass filter is applied to eliminate the effects of biorhythms such as respiration or heart rate. The cut off of this filter is typically 128 seconds. Due to the serial acquisition of the fMRI data time-series successive time points will be correlated. To account for this temporal auto-correlation an autoregressive model of order 1 + white noise is fitted to the data. The β parameters (often referred to simply as 'betas') for each voxel are then estimated by multiple

linear regression so that the sum of the squared differences between the observed data and the values predicted by the model is minimized.

### II.3.2 T and F-statistics

Inferences about the relative contribution of each explanatory variable (x), each represented by one column in the design matrix, can be made by conducting T or F-tests on the parameter estimates. The null hypothesis that the parameter estimates are zero is tested by an F-statistic, resulting in an SPM(F). To compare the relative contribution of one explanatory variable compared to another one can contrast or subtract the parameter estimates from one another, and test whether the result is zero using a t-statistic, resulting in an SPM(t). The t-statistic is calculated by dividing the contrast of the parameter estimates by the standard error of that contrast. To make inferences about regionally specific effects the SPM(t) or SPM(F) is thresholded using height and spatial extent thresholds specified by the user.

## II.4 Visual Area Localization

The human visual cortex consists of multiple subregions: primary visual cortex V1 (also sometimes referred to as striate cortex) and

extrastriate regions V2 and V3 as well as a number of higher level visual areas. The response properties of neurons contained in these regions are often well studied and, in most cases, differ drastically from area to area. Precise delineation of the borders of early visual areas is of crucial importance to the success of fMRI studies of human vision. In addition, it is often necessary to localize activations in the occipital cortex to specific early visual areas. However, there is wide inter-subject anatomical variability of early visual areas which precludes the assignment of visual area borders based on stereotactically normalised coordinates (Dougherty et al., 2003). Thus, unless a more accurate method is used, voxels representing adjacent visual areas (with very different neuronal response properties) will often be incorporated into a single 'visual' region of interest or visual activations will be mislocalized. Fortunately, early visual cortical areas are retinotopically organized, that is, their neurons respond to stimulation of limited receptive fields whose centers are organized to form a continuous mapping between the cortical surface and the contralateral visual field. This consistent organization can be utilized to accurately determine the boundaries between early visual cortical areas using fMRI (Engel et al., 1994; Sereno et al., 1995).

### II.4.1 Anatomy of visual areas

Within each hemisphere, human area V1 occupies a roughly 4- by 8-cm area located at the posterior pole of the brain in the occipital lobe. A large fraction of area V1 falls in the calcarine sulcus. From

posterior to anterior cortex, the visual field representation shifts from the centre (fovea) to the periphery. The midline of V1 represents the horizontal meridian, while the boundary of V1 and V2 represents the vertical meridian (both dorsally and ventrally). The local representation of the visual field on the cortical surface changes its orientation at the boundaries between V1 and V2 (and V2 and V3). Therefore, the spatial extent of activations elicited by visual stimuli representing the horizontal and vertical meridians can be used to functionally define these borders. This technique is called meridian mapping, and is a rapid method of retinotopic mapping.

Figure II-1 illustrates meridian mapping. Participants were stimulated with flashing checkerboards along the vertical meridian (VM) or the horizontal meridian (HM, Figure II-1A). This produced a differential activation profile around the occipital pole: "stripes" of activity related to either the vertical or horizontal meridian can be observed in either an inflated or flattened view. Figure II-1B depicts such activity overlaid onto an inflated right hemisphere occipital pole (medial-occipital view). Once the visual borders are delineated manually the so-defined visual areas V1, V2 and V3 can be overlaid onto an uninflated view of the same cortex (gray matter view, Figure II-1C). While meridian mapping is very fast, it provides poor information about eccentricity encoding within visual areas, and is not typically able to accurately define V4. To overcome these limitations usually requires the use of phase encoded retinotopic mapping methods (using a rotating wedge and expanding ring stimulus to generate a spatiotemporal pattern of stimulation of the visual field). In this thesis meridian mapping was used in all studies

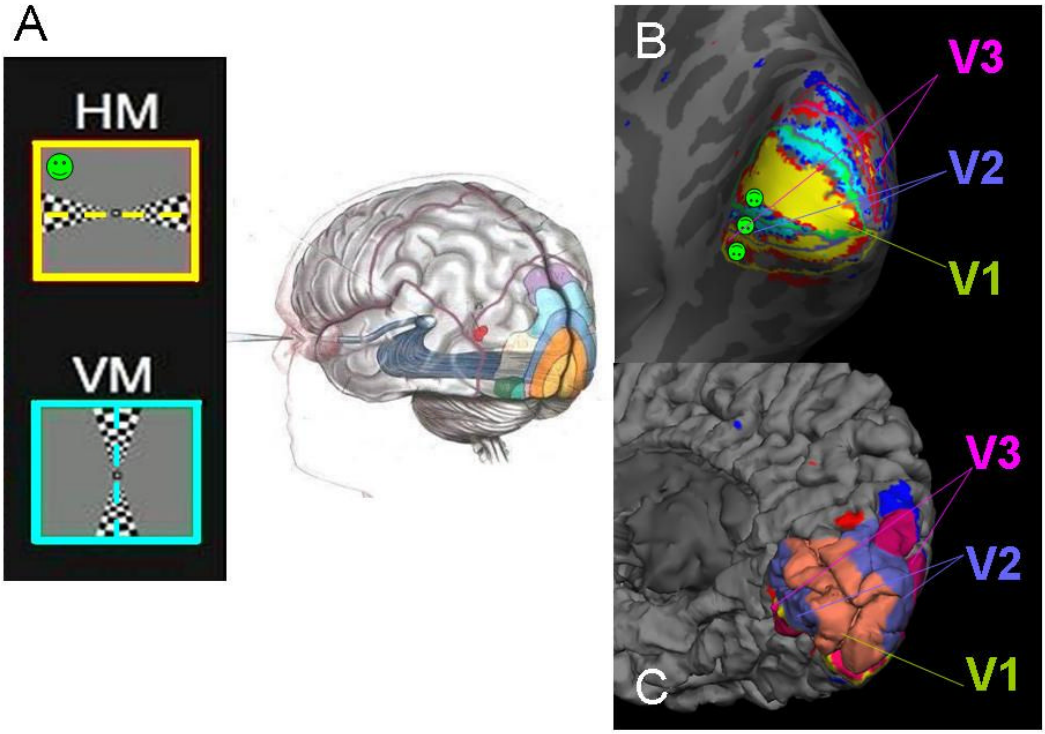



**Figure II-1**
**Meridian mapping procedure illustrated on an actual participants brain (S1, Chapter 5). A: Stimulation with flashing checkerboards along the vertical meridian (VM) or the horizontal meridian (HM) leads to a differential activation profile in the occipital pole. B: Overlaid onto an inflated 3D-reconstruction of the right hemisphere (medial-occipital view) are activity peaks for stimulation from A –HM activity in yellow and VM activity in cyan. The horizontal stimulus activated the midpoint of the calcarine sulcus and the vertical stimulus the gyri on either side of it. This alternating pattern of activation by horizontal and vertical meridian stimulation can be used to map the boundaries of early visual areas. The borders of visual cortices V1, V2 and V3 are displayed in green, purple and pink. Additionally the orientation and position of a fictive 'smily' illustrates position and eccentricity distribution in these retinotopic cortical areas. C: Once mapped, the overlay of the regions onto a white-matter-3D plot illustrates how V1 lies along the calcerine sulcus, flanked by ventral and dorsal parts of V2 and V3.**

as the relevant experimental questions did not require accurate eccentricity information and were limited to V1 to V3.

A standard protocol to identify the boundaries V1, V2 and V3, includes flickering checkerboard patterns, displayed either at the horizontal or vertical meridian, alternated with rest periods for about 10-20 epochs of about 15-30s over two sessions. Delineating

the borders between visual areas using activation patterns from this meridian localizers (illustrated in figure II-1) results in mask volumes for each region of interest (left and right V1, V2 dorsal, V2 ventral, V3d, and V3v).

## II.5 Multivariate Pattern Recognition for fMRI

In recent years, multivariate pattern decoding (MVPD) has proven itself as a powerful tool in the analysis of fMRI data. By taking into account subtle biases in the pattern of activity recorded, measured simultaneously at many locations, subtle biases in the pattern of activity have been demonstrated to allow the study of processes of the human brain thought to be represented below the spatial resolution of fMRI (Haxby et al., 2001; Cox and Savoy, 2003; Mitchell et al., 2003; Haynes and Rees, 2005b, 2005a; Kamitani and Tong, 2005b, 2005a; Haynes and Rees, 2006; Kamitani and Tong, 2006; Kriegeskorte et al., 2006; Serences and Yantis, 2006; Haynes et al., 2007; Serences and Boynton, 2007a; Formisano et al., 2008; Mitchell et al., 2008; Sumner et al., 2008). As a direct consequence, it is now generally deemed promising to study neural representations that have been previously thought inaccessible to non-invasive imaging techniques in humans.

As an essential part of this thesis, I developed and implemented a set of MatLab functions that form a toolbox for MVPD. The toolbox consists of two basic parts: pre-analysis and classification. In the process of developing this MVPD-toolbox I have tried multiple

algorithms to actually perform classification starting with simple Linear Discriminant Analysis (LDA) to more sophisticated Support-Vector-Machines (SVM), but I finally settled on spare logistic regression (SLR, Yamashita et al., 2008). SLR has the invaluable advantage to have a built-in procedure for voxel selection that makes voxel preselection, beyond predefining specific areas of the brain spatially, unnecessary. In the following I briefly describe the components of the two basic parts of the toolbox; for the actual MatLab-functions please refer to the appendix.

### II.5.1 Pre-analysis for MVPD

Before actually classifying two or more different conditions of an fMRI experiment from each other, a few simple steps have to be performed. However, after realignment these differ from conventional analysis are thus explained in brief.

Firstly, the hemodynamic delay of the BOLD signal (see above) has to be adjusted for. In general, this can be either be done by convolving the entire timecourse of each participants scans with an HRF, or by convolving the condition onsets. In this thesis I have always applied the latter using a simple step-function to approximate the time-delay in the BOLD signal. The exact timing of this step-function was dependent on each study's individual TR, but was usually between 6-8 seconds.

Secondly, a critique of early MVPD studies was that it might be possible for voxels to be autocorrelated across entire sessions due to technical details of the image reconstruction of fMRI in general. To avoid such problems 'leave-one-out cross-validation' has become the standard method to evaluate classification accuracy (Pereira et al., 2009). Thus, in all experiments of this thesis I have applied a leave-one-out cross-validation between all sessions recorded.

Thirdly, in order to evaluate the probability that classification is driven by over-fitting of arbitrary patterns of spatial correlations in the data that have nothing to do with the conditions analyzed, a shuffle-control test was carried out for all experiments (Mur et al., 2009). In this test the assumption that classification is driven by chance is tested: if it were true, similar results should be obtained if labels indicating the experimental condition for each example vector were shuffled randomly. To test this, a separate analysis is performed in which labels of the test examples are re-shuffled for each round of the cross-validation procedure for each experiment of this thesis. The resulting distribution of classification accuracy characterizes the expected distribution of accuracy under null hypothesis.

Fourth, as a general rule for any classical MVPD algorithms, results become less reliable when the number of input-features is higher than the number of input-vectors. This phenomenon is sometimes described as overfitting (Cristianini, 2000). In fact, some influential implementations of classification algorithms (e.g. the LDA-function "classify" in MatLab) do not accept such unbalanced pairs at all. However, others algorithms (e.g. SVM) are said to be relatively

robust against this problem as long as the imbalance is not too skewed (Cristianini, 2000; Carl, 2004). In the context of MVPD for fMRI data the number of input vectors is determined by the number of volumes per condition. For example, in a study comparing 2 conditions, where each condition appeared once per session for 8 volumes and 10 repetitions of sessions are recorded, this means that there are 80 vectors per condition. Due to the leave-one-out procedure this would result in 10 pairs of 144 training and 16 test vectors. Thus, in this simple example, it would not be advantageous to consider more than 144 input features. The input features that form fMRI data, however, are voxels. An entire brain volume at a resolution of $1.5mm^2$ inplane and about 30-40 slices per volume can have more than 500000 voxels. Thus, before classification, the number of voxels to consider has to be dramatically reduced. In general, this is mainly achieved by limiting classification to certain regions of interest (ROI) in the brain. These ROIs can be either spatially or functionally defined (e.g. by meridian mapping), however, ROI definition has to be independent of the main experiment (Kriegeskorte et al., 2009). However, even small ROIs often consist of several hundreds to thousands of voxels which is why in general a further voxel-selection becomes necessary. This further voxel selection is often achieved by thresholding all voxels in an ROI by an experimental contrast. However yet again, to avoid circularity in the analysis, this contrast must be independent of the experimental conditions tested (Kriegeskorte et al., 2009). For this reason, MVPD studies using LDA or SVM often either ignore this final selection step (and thus risk overfitting) or use an irrelevant contrast for thresholding. Another common method is to sort voxels

on the basis of independent criteria, however it is truly important that these sorting criteria are truly independent – a fact that was dramatically demonstrated by a recent study and subsequent comments on its non-independent sorting criteria amongst other methodological errors (Grill-Spector et al., 2006; Baker et al., 2007; Simmons et al., 2007). In this thesis, all MVPD results are obtained by with an SLR algorithm (Yamashita et al., 2008). Different to manual sorting or thresholding, SLR employs a unique method termed automatic relevance detection (ARD) for voxel selection (see below).

Finally, before classification, vectors that are classified are generally normalized. The individual options for normalization vary greatly from study to study: Normalization generally entails either z-scoring or simply mean-correcting per scan, either all extracted vectors per scan or each scan as an entity. The difference is that in the first case only vectors actually used for classification influence the normalization, while in the second case other information might influence the normalization (e.g. baseline-blocks). Depending on the experimental paradigm both outcomes might be desirable. However, for experiments in this thesis, only vectors actually used during classification have been normalized by z-scoring.

### II.5.2 SLR and ARD

SLR is a Bayesian extension of logistic regression that combines an innovative strategy for adaptive, yet unbiased voxel selection with

conventional linear discriminant analysis. Conversely to sub-optimal solutions like sorting or thresholding, SLR selects voxels solely on the basis of the training-set for each cross-validation, a process termed automatic relevance detection (ARD) which is guaranteed independent of the test data. Yamashita et al (2008) first presented SLR for fMRI multivariate analysis, which contains ARD as an essential part, and demonstrated the superiority of the SLR algorithm when compared to other MVPD algorithms (Yamashita et al., 2008). In this thesis, all MVPD results are obtained with the SLR algorithm as presented by Yamashita and colleagues.

During ARD, within every iteration of the cross-validation, SLR carries out a number of *nested cross-validations* inside the training set: the training set is divided randomly in two sections of specified proportion; for a randomly selected subset of the voxels, the linear classifier is trained with one section of the data and tested with the other and the selected voxels are weighted proportional to the accuracy of this classification. This procedure is carried out 500 times while voxels accumulate weights. By the end of the nested cross-validation, the assigned weight of each voxel is taken as a relevance factor indicating how informative the voxel is for classification. Voxels with the highest relevance are then selected for the actual classification of test data. Importantly, this voxel selection algorithm depends entirely on the training set and is completely ignorant about and independent of the test set. Once the voxels are selected, the training and test data from the selected voxels are then passed on to a conventional linear classifier (Yamashita et al., 2008).

In all studies of this thesis, classification accuracy is averaged across the cross-validations for each ROI in each participant. Then, statistical significance of the results is evaluated across all participants for each ROI using a student's t-test. To adjust for multiple comparisons due to multiple ROIs, these results are then Bonferroni-corrected.

## II.6 Conclusion

This chapter has described conventional and multivariate fMRI analysis and visual area localization, the methods that were used in all of the experiments presented in this thesis. I have presented a summary of the physics of fMRI, the physiological basis of the BOLD signal and the statistical basis of SPM. In addition, I have discussed the physiological basis of retinotopic mapping. However, for practical reasons the precise use of these methods sometimes varied across experiments and each experiment utilized additional methods. Therefore each experimental chapter in this thesis has a methods section describing these points in more detail.

# Chapter III:

# Effects of attentional load on orientation selective processing

## III.1 Introduction

Attentional load describes a selection mechanism that depends on the availability of processing resources. Whereas low attentional load during an attended, goal-directed, task leaves spare attentional capacity to 'spill over' to process goal-irrelevant distractor stimuli, high attentional load takes up all capacity and thus decreases distractor processing (Lavie et al., 2004). Numerous experimental paradigms have empirically tested the influence of attentional load. Experiments commonly manipulate attentional load in a primary task and then measure the influence of this attentional load manipulation on task-irrelevant distractor stimuli. Studying paradigms like this, high attentional load has been found to severely reduce task-irrelevant, distractor related behavioral measures (Lavie, 1995; Lavie and Fox, 2000), as well as to reduce task-irrelevant, distractor related neural activity (Rees et al., 1997; O'Connor et al., 2002; Pessoa et al., 2002; Pinsk et al., 2004; Schwartz et al., 2005). One of two possible results are generally reported in such studies of distractor related neural activity: Either a simple main effect of reduced distractor activity under high load

(i.e. Schwartz et al., 2005 for checkerboard representations in early visual cortices), or an interaction effect between the distractor presence (vs. absence) and low (vs. high) load (i.e. Rees et al., 1997 for activity to motion vs. static dots in V5/MT). However, in both cases it remains unclear what the reduced activity (the reduced difference in activity) in a brain area actually means for the underlying representation of the distractor stimuli. It could be that the distractor related activity underwent modulations in neural firing rate, tuning curves of individual neurons could be modulated or the area in question could have received less feedback input. Thus, the question 'what aspects of activity are modulated' remained elusive.

### III.1.1 Objectives

With the experiment presented in this chapter, I investigate one specific aspect of distractor related activity under attentional load manipulation: basic visual feature-specific processing. Basic visual features, like colour, orientation or direction of motion, are some of the best studied neuronal architectures in the brain (Mountcastle, 1997) with most of these features represented in a columnar architecture in early visual cortices. For orientation, for example, neurons with different orientation preferences are systematically mapped across the cortical surface of primary visual cortex, with regions containing neurons of similar orientation preferences separated by approximately 500 µm (Hubel and Wiesel, 1962, 1968; Wang et al., 1996). For orientation amongst other visual features, it has recently emerged that fMRI can be used to study

fine-grained neural representations or different orientations, even though they are encoded at a finer scale than the resolutions of fMRI (Haynes and Rees, 2005a; Kamitani and Tong, 2005b; Haynes and Rees, 2006). These (orientation-) specific studies employed multivariate pattern decoding (MVPD), a technique allowing the study of orientation processing in the human brain by taking into account subtle biases in the spatial pattern of activity (see Chapter I and II for a detailed description of multivariate pattern decoding and its application in fMRI studies).

In this experiment, I sought to investigate whether orientation-specific distractor processing depended on attentional demands in an unrelated task. I report results from early visual cortices V1, V2 and V3. To measure the orientation-specific signal associated with (distractor related) brain activation, the accuracy of multivariate pattern classification of brain activation patterns was used as an indicator of the quality of the representation of orientation. This experiment thus sought to establish whether and how the feature selective processing of orientation direction in early visual cortex is modulated by attentional load.

### III.1.2 Hypothesis

Here, I combined elements of experimental designs previously investigated. The distractor stimuli were identical to Haynes & Rees (2005); the central load task is identical to Schwartz et al (2005). Accordingly, my first hypothesis was to replicate behavioral effects

of varying load for the central task as previously reported. Additionally, I also sought to replicate the univariate main effect of reduced brain activity for the peripheral distractor in early visual cortices (Schwartz et al., 2005).

Secondly, I sought to replicate successful multivariate pattern decoding of neural representations of oriented stimuli (Haynes and Rees, 2005a). Extending the general predictions of load theory to feature-specific processing, I hypothesized that there would be a reduced representation of orientation-specific signals (as measured by multivariate classification accuracy) under high load (compared to low load). This result would extend load theory validity to feature-specific neural processing. Fourth, I predicted that the multivariate decoding of load (high vs. low rather than left vs. right) would be significantly less successful than the classification of orientation – potentially even not significantly different for chance. This is due to the assumption that multivariate pattern decoding of visual features is likely to rely on biased voxels whose signals reflect activity of neurons in early visual areas tuned to elementary features such as colour and orientation. Since load is not such an elementary feature it should classify significantly worse or not at all.

Finally, as a fifth hypothesis, I predicted that it would be possible to generalize from one set of orientation training data (under one load condition) to a second orientation data set taken from the other experimental condition (the other load condition). I intended to test this generalization on training with low, and testing with high, data and vice versa. My hypothesis was that it would be possible to

generalize from low to high load and vice versa, as (orientation-) biased voxels should remain the same across conditions.

## III.2 Methods

### III.2.1 Participants

Eight healthy participants (4 male, mean age 24.1 years) gave written informed consent to participate in the study, which was approved by the local ethics committee. All subjects had normal or corrected to normal vision.

### III.2.2 Stimulus

During the main fMRI experiment, participants performed a visual detection task on a continuous rapid successive visual presentation (RSVP) of coloured crosses (one cross every 750 ms) that was shown in a fixed central location at fixation. This RSVP stream consisted of cross-shaped stimuli with two different orientations (upright or upside-down) and six different colours in random order (Figure III-1 B). Participants were required to monitor the occurrence of infrequent (7.5%) pre-specified targets within this rapid central letter stream, and to respond by a button-press to each detected target. Behavioural responses were collected with a standard MRI-button box.

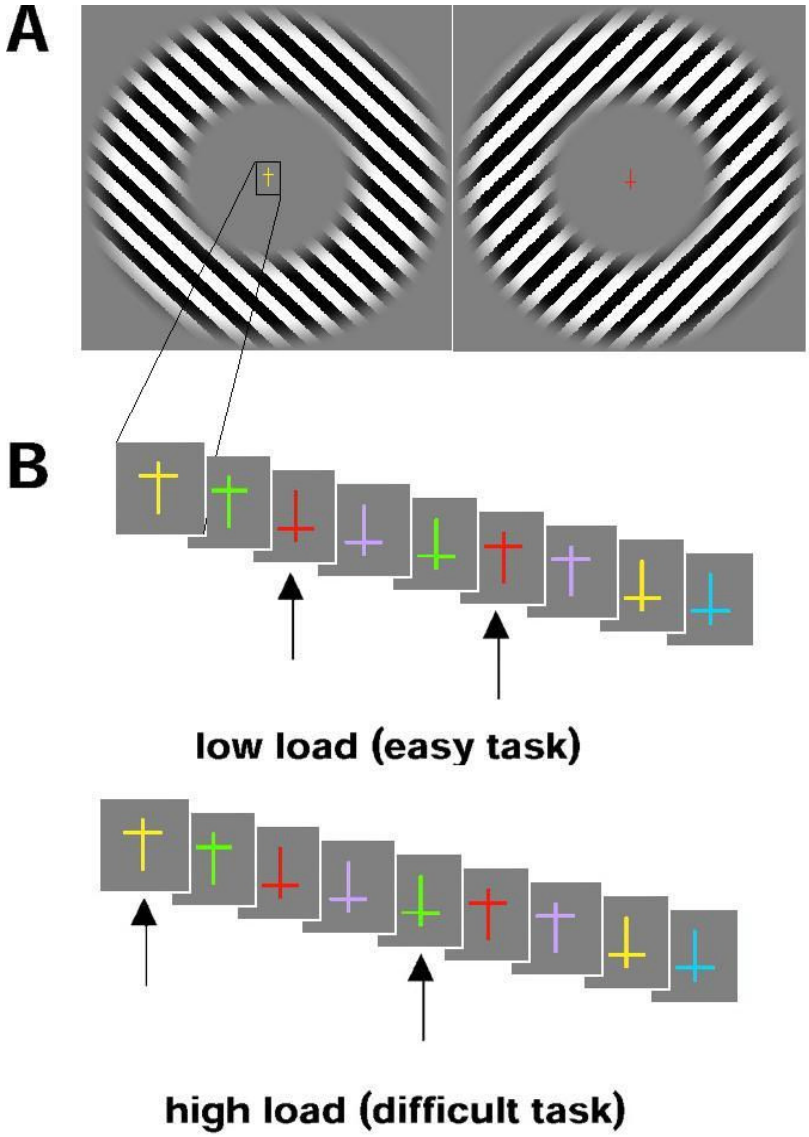



**Figure III-1**
**Example stimuli used in the experiment.**
**A) Shown are two examples of stimuli used in the main experiment (during the experiment only one stimulus was shown at a time). Stimuli consisted of oriented lines that were continuously contrast reversed at a frequency of 4Hz. Orientated lines comprised an annulus shaped field with a radius between 4° and 8°. At fixation the load task is performed by the subject.**

**B) The load task consisted of a visual detection task on a continuous rapid successive visual presentation (RSVP) of colored crosses (one cross every 750 ms). Subjects were required to detect infrequent (7.5%) pre-specified targets: red crosses of either orientation (low load, easy task) or a combination of yellow upright or inverted green crosses (high load, difficult task).**

The low-load (colour) task required a key-press for any red cross irrespective of its orientation whereas the high-load (conjunction) task required a key-press for any upright yellow cross or upside-down green cross. Items that were targets in one condition also appeared with the same frequency as task-irrelevant stimuli in the other condition (i.e. high-load targets appeared as distractors under low-load instructions, or vice versa). Consequently, only the task instructions distinguished the high-load and low-load conditions for the central task. The rapid succession of stimuli and unpredictability of targets in this RSVP task ensured that participants always monitored items at central fixation, during both task conditions.

Based on prior, identical usage of this task at the centre of gaze (Schwartz *et al.*, 2005) I anticipated that detecting red targets would be a low-load task that can be solved on the basis of a single 'pop-out' colour feature, whereas monitoring for crosses with a particular colour and orientation in the rapid central stream should be a high-load task requiring high attentional resources in order to discriminate the specific conjunction of features. In addition to the central task, visual stimuli comprised a ring shaped annulus with a radius between 4° and 8°, made up of alternating black/white diagonal stripes reversing contrast at 4Hz Figure III-1 A).

### III.2.3 Procedure

Participants lay supine in the scanner and viewed visual stimuli that were projected from an LCD projector (NEC LT158, refresh rate 60

Hz) onto a screen viewed via a mirror positioned within the MR head coil. Stimuli were presented using MATLAB (Mathworks Inc.) and COGENT 2000 (www.vislab.ucl.ac.uk/Cogent/index.html). Complete darkness was achieved in the scanning environment by manually masking the fMRI projection screen, head coil and internal bore with matt black card. This eliminated discernable non-retinotopic luminance cues, ensuring that the only source of visual stimulation during experimental runs was the experimental stimulus.

All participants were scanned for a total of 4 to 6 runs, alternating the orientation and difficulty of the central task in a manner that was counterbalanced within each run and across participants. Each block started with the appearance of the central fixation spot for 10s including a short reminder of the task to come for 3 seconds, followed by a 30s long lasting block of the load task at central fixation with the flickering annulus surrounding the central task. The last 10s of each block were used to give participants a short rest and only comprised a fixation cross. Participants were instructed to concentrate on the central task/ fixation cross for the duration of the entire run. Between each run the screen was masked to prevent scattered light from the projector illuminating it.

### III.2.4 Scanning

The main experimental task lasted 400s split into 8 parts of 50 seconds each. Hence, each experimental fMRI scan comprised each of the four task conditions twice (low/ high load combined with left/

right oriented lines), resulting in 8 trials per scan. Additionally to the experimental sessions, to identify the boundaries of primary visual cortex (V1) and extra-striate retinotopic cortical areas V2 and V3, standard retinotopic mapping stimuli were presented twice to each participant. During these runs participants solely have to fixate on a central fixation cross. Retinotopic mapping runs lasted for 165 image volumes. In total, 6 to 8 scanning runs of 165 to 310 image volumes were acquired per participant.

*Imaging Parameters.* A 3T Siemens Allegra system acquired T2*-weighted Blood Oxygenation Level Dependent (BOLD) contrast image volumes using a descending sequence every 1.3s. Each volume comprised 20 3-mm-thick slices, positioned on a per participant basis to give coverage of the occipital lobe with an in-plane resolution of 3x3 mm. To maximize signal to noise in early visual cortex an occipital head coil was used.

### III.2.5 Analysis

*Data preprocessing.* Data were preprocessed using Statistical Parametric Mapping software (SPM5, www.fil.ion.ucl.ac.uk/spm). After discarding the first seven image volumes from each run to allow for T1 equilibration effects, functional image volumes were realigned to the first of the remaining volumes and co-registered to the individual participants' structural scans. Experimental data and retinotopic mapping data were not spatially smoothed. Data were

high-pass filtered (cut-off – 128s) to remove low-frequency signal drifts.

*Visual Area Localization.* To identify the boundaries of primary visual cortex (V1) and extra-striate retinotopic cortical areas V2 and V3, standard retinotopic mapping procedures were used (Sereno *et al.*, 1995). Checkerboard patterns, displayed either at the horizontal or vertical meridian, were alternated with rest periods for 16 epochs of 26s over two scanning runs. Mask volumes for each region of interest (left and right V1, V2 dorsal, V2 ventral, V3d, and V3v) were obtained by delineating the borders between visual areas using activation patterns from the meridian localisers. I followed standard segmentation and cortical flattening in MrGray (Teo *et al.*, 1997; Wandell *et al.*, 2000) to determine the borders between the ROIs on a flattened cortical representation.

*Univariate analysis.* Initially, I used a standard univariate approach to determine whether there were any differences in activation comparing the different types of distractor (left or right tilted) and different levels of attentional load in the central task (low or high load). This was achieved by creating a standard SPM5 univariate first level analysis (see Chapter II: General Methods) defining 4 box-car regressors for the 4 experimental conditions. The resulting beta-images, one for each box-car regressor, represent brain activity effect size at each particular voxel in percentage units of the whole brain mean, convolved by the standard hemodynamic response (HDR) function. The standard approach is to then formulate a contrast within SPM. However, in order to compare the

4 experimental conditions precisely per ROI, I calculated the mean of all voxels of all beta-images within a given ROI of each condition. This procedure allowed me to reduce the entire dataset acquired to 4 values, one for each experimental condition (Figure III-3), per ROI, per participant.

In detail, data were collected in 4 scans of 4 blocks for each subject, blocks were separated by rest. Each scan contained each condition exactly once in pseudo-randomized order, fulfilling a completely balanced design. For each subject we first averaged over all voxels in a ROI, in all volumes of a block. Then we averaged these block-values across scans, resulting in 4 values of mean brain activity per subject.

Due to the necessity that all 4 values are relative to the same mean brain activity per subject, a repeated measures ANOVA, where each subject is a new repetition, was used to compare for differences of brain activity between conditions.

Finally, I compared high vs. low load brain activity, collapsing across orientation. This represents a simplified version of the above procedure, reducing the entire dataset into two groups rather than four. Pairs of conditions were always defined as values from the same subject and, thus, tested with a paired t-test.

*Multivariate pattern classification.* Pattern classification was performed using Sparse Logistic Regression (SLR, see Chapter II: General Methods). For determining classification accuracy, only

classification with unseen and independent-test data was considered. Thus, test data sets in different iterations were always independent from the training data sets used. The actual classification was repeated n times, where n was determined by the number of independent blocks (n-fold-cross-validation). Classification accuracies were averaged across these n data assignments. Note that these n iterations are completely independent of each other and there was no iterative learning or similar techniques applied. Much care was taken to ensure independence of training and testing vectors at all times.

Multivariate pattern classification was initially used to predict the orientation of the irrelevant flickering distractor annulus. This left vs. right orientation classification was repeated for low and high load, strictly keeping data from the two conditions separate, resulting in separate results for MVPD under low and high load. We then tested for a significant difference from chance (50%) with a simple Student's t-test independent for low and high load.

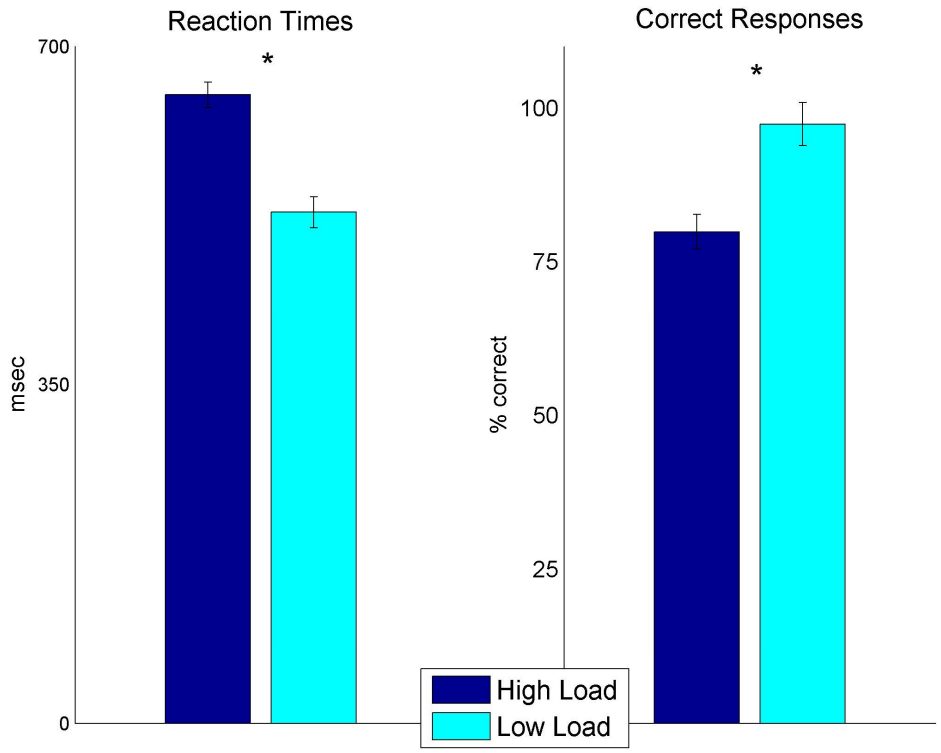



**Figure III-2**
**Behavioral Results. The left panel shows mean reaction times averaged across participants for successful detection of target crosses.**
**Separate colors denote low (blue) and high (red) load conditions during the central RSVP task (see fig 1).**
**The right panel shows mean correct hits across participants for target crosses.**
**Both differences are highly significant and present in every participantupright or inverted green crosses (high load, difficult task).**

After predicting orientation (left vs. right), MVPD was repeated, now attempting to classify load (high load vs. low load). In other words, instead of classifying left vs. right tilted orientations in one load condition, I attempted to classify left vs. left (and right vs. right) tilted orientations in different load conditions. Significant difference form chance (t-test, as above) in this classification step was thus a test of load decoding in early visual cortex.

Finally, I repeated the entire multivariate analysis to test for the ability to generalize from feature classification in one load condition to test data taken from a different experimental condition. This generalization decoding performance was computed using training on data from low load orientation data (left vs. right under low load) and testing on high load orientation data (left vs. right under high load), or vice versa. The same was repeated for the second part of the MVPD analysis, attempting to generalize from left vs. left training (different load conditions) to right vs. right-test data, and vice versa.

## III.3 Results

### III.3.1 Behavioral results

For both tasks, performance was always above chance, confirming adherence to the task requirements. The central task was significantly harder for the high-load than low-load condition in all participants. I initially examined Reaction Times (RT) across participants comparing the RTs of low load condition with the RTs of high-load condition with a paired, two-tailed t-test. Mean detection latencies for central target crosses were significantly slower in the high-load versus low-load condition [mean Reaction Time high: 650ms; low: 528ms; $t(25) = 6.8$, $P < 0.001$] (Figure 1-3, right).

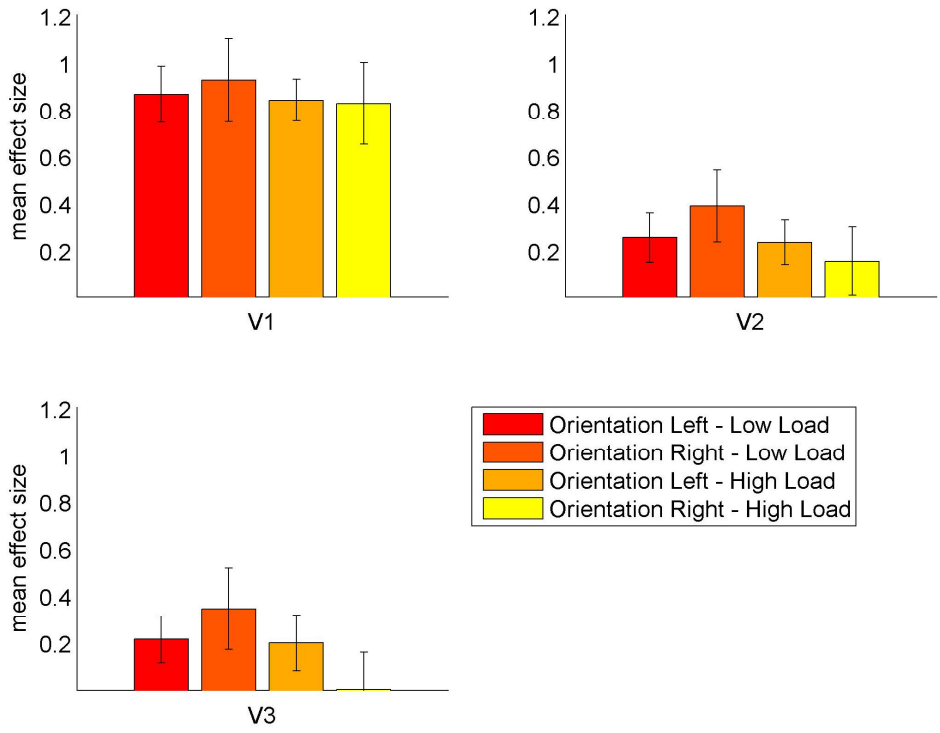



**Figure III-3**
**Univariate analysis. Shown is the mean univariate effect size across subjects under the four different experimental conditions (colors). Mean effect sizes are obtained from beta images for each condition separately from unsmoothed data. Errorbars denote SEM across subjects. An ANOVA revealed no significant differences between any of the conditions and no significant interactions.**

I also analyzed hit rates of correctly detected crosses in both conditions. Again I discovered a significant difference between high and low load condition. Hit rates significantly decreased in the high load condition [mean Hit Rate high: 79.8%; low: 97.4%; $t(25) = -15.4$, $P < 0.001$] (figure 1-3, left).

Finally, I examined the error rates (incorrect target responses or false alarms, not including missed responses) using a similar paired t-test. I found a significant difference between the two conditions. False alarms significantly increased during the high load task [mean error rate high: 31.3%; low: 16.6%; $t(25) = 8.1$, $P < 0.001$].

Performance was always reliably above chance on either task, confirming adherence to the task requirements. Taken together, this behavioral data confirms that participants successfully paid attention to the central task and that central attentional load was effectively varied by my task manipulation.

### III.3.2 Univariate Results

I analyzed the mean univariate effect sizes across participants under the four different experimental conditions (Figure III-3). A repeated measurements (RM-) ANOVA revealed no significant main effects or interactions between the mean signal of the representation of the task-irrelevant annulus V1. However, in V2 and V3 there was significantly lower activity under high (vs. low) load. In both V2 and V3 there was also a significant interaction of load and orientation. [RM-ANOVA: V1Load: $F(1,7) = 1.5$, $p = 0.26$; Orientation: $F(1,7) = 0.0334$, $p = 0.86$; Load x Orientation: $F(1,7) = 0.486$, $p = 0.508$; V2: Load: $F(1,7) = 10$, $p = 0.0157$; Orientation: $F(1,7) = 0.0762$, $p = 0.791$; Load x Orientation: $F(1,7) = 9.64$, $p = 0.0172$; V3: Load: $F(1,7) = 10.1$, $p = 0.0156$; Orientation: $F(1,7) = 0.146$, $p = 0.714$; Load x Orientation: $F(1,7) = 15$, $p = 0.00609$].

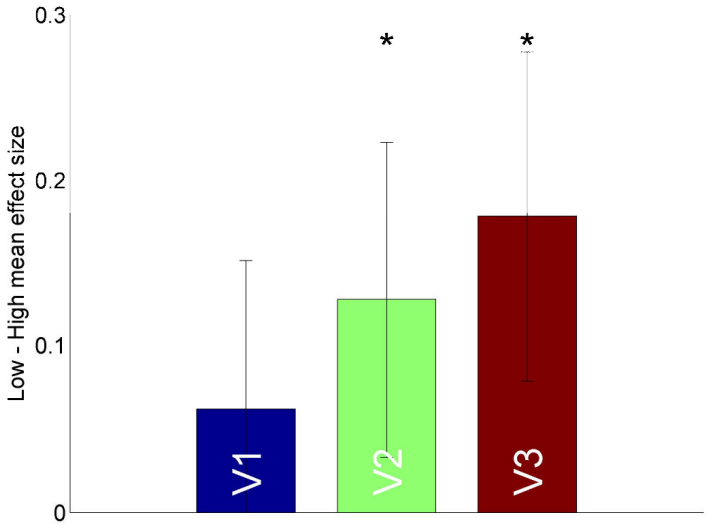



**Figure III-4**
**Univariate Low mean effect size minus High mean effect size. Similar to Figure 3 values are obtained from beta images, however here the mean of all values collected under high load is subtracted from the mean of values collected under low load, collapsing across orientations. A ttest revealed significantly reduced activity under high load for V2 and V3, replicating previous BOLD activity findings at similar eccentricity of the distractor (Schwartz et al., 2005, Fig 7 and 8).**

Next, I calculated the main effect of load again separately, collapsing across the orientation of the distractor annulus. Figure III-4 shows the difference between the mean activity under high load from the mean under low load. The underlying values were then compared using a paired t-test for significant differences between low and high load pairs (pair defined as from the same scan-run). The result confirmed the ANOVA findings: mean effect size of BOLD signal was significantly higher under low load than under high load in early visual cortices V2 and V3 [V1: Low load: 0.9, High load: 0.84, t(15) = 1.2, P = 0.24402; V2: Low load: 0.33, High load: 0.2, t(15) = 2.8, P = 0.013276; V3: Low load: 0.29, High load: 0.11, t(15) = 2.8, P = 0.013826].

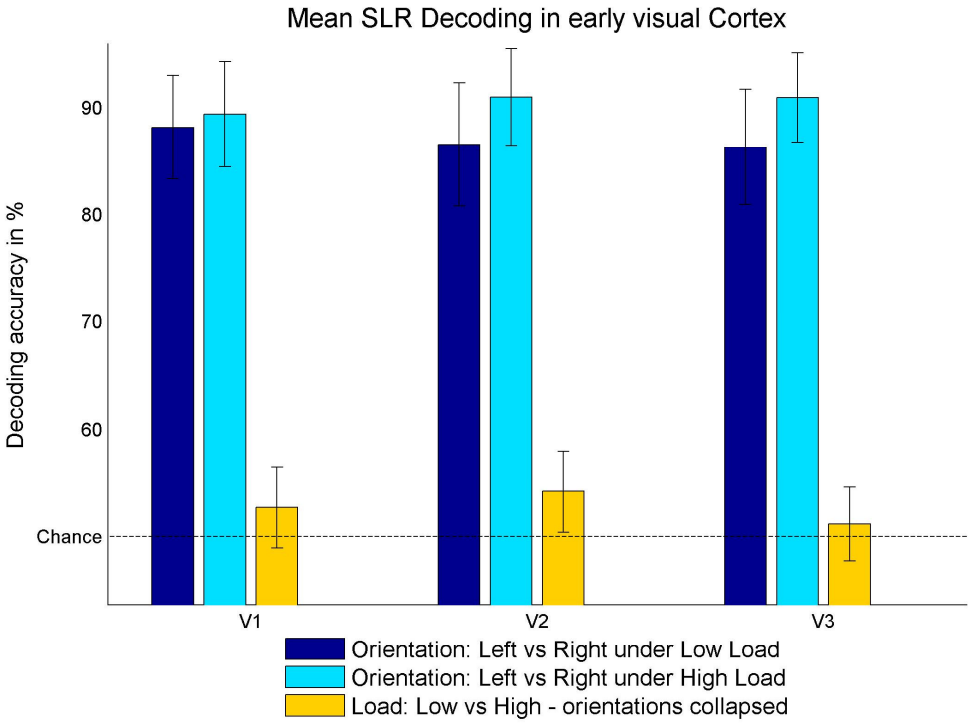



**Figure III-5**
**Decoding accuracy (mean across subjects) is shown for visual areas V1 to V3, obtained with Sparse Logistic Regression (SLR, see General methods).**
**Different colors indicate decoding for different comparison: leftward vs. rightward orientation decoding was highly successful, compared to chance, under both low load (dark blue bars) and high load (light blue bars) conditions. However, ttests between the two conditions revealed no significant differences in any area. I then attempted to decode Load on trials from one orientation only. This comparison, however, proved non-significantly different from chance (yellow bars, collapsed across rightwards only and leftwards only decoding). Errorbars denote SEM across subjects.**

### III.3.3 Multivariate classification results

Figure III-4 shows the decoding accuracy (mean across participants) for visual areas V1 to V3, obtained with Sparse Logistic Regression (SLR). The y-axis depicts mean accuracy of classification across participants, while different colors along the x-axis indicate decoding from different experimental conditions in visual areas V1

to V3. Decoding leftward vs. rightward orientation was highly successful, compared to chance, under both low load and high load [V1: Low Load: 88.2%, $t(7) = 11$, $P < 0.01$; High Load: 89.4%, $t(7) = 15$, $P < 0.01$; V2: Low Load: 86.6%, $t(7) = 11$, $P < 0.01$; High Load: 91%, $t(7) = 13$, $P < 0.01$; V3: Low Load: 86.3%, $t(7) = 9.4$, $P < 0.01$; High Load: 90.9%, $t(7) = 13$, $P < 0.01$].

Next, I attempted to decode Load (low vs. high) on blocks from one orientation only. This comparison, however, proved substantially less successful compared to orientation classification, with prediction accuracy in all cases and ROIs very close to chance. However, although SLR decoding revealed a weak result across participants, it was still significantly different from chance in V1 and V2 for an uncorrected t-test, classifying low load vs. high load on data from one orientation only [same orientation (collapsed): V1: 52.7%, $t(15) = 2.3$, $P = 0.038$; V2: 54.2%, $t(15) = 2.8$, $P = 0.013$; V3: 51.2%, $t(15) = 0.76$, $P = 0.46$].

Next, I compared low and high load condition results across participants using uncorrected t-tests to determine whether there were significant differences in classification accuracy comparing different attentional load conditions. These t-tests between these two conditions revealed no significant differences in any area [t-test, Low Load 88.2%, High Load 89.4%, $t(7) = -0.57$, $P = 0.59$; V2: Low Load 86.6%, High Load 91%, $t(7) = -1.9$, $P = 0.11$; V3: Low Load 86.3%, High Load 90.9%, $t(7) = -2.2$, $P = 0.07$].

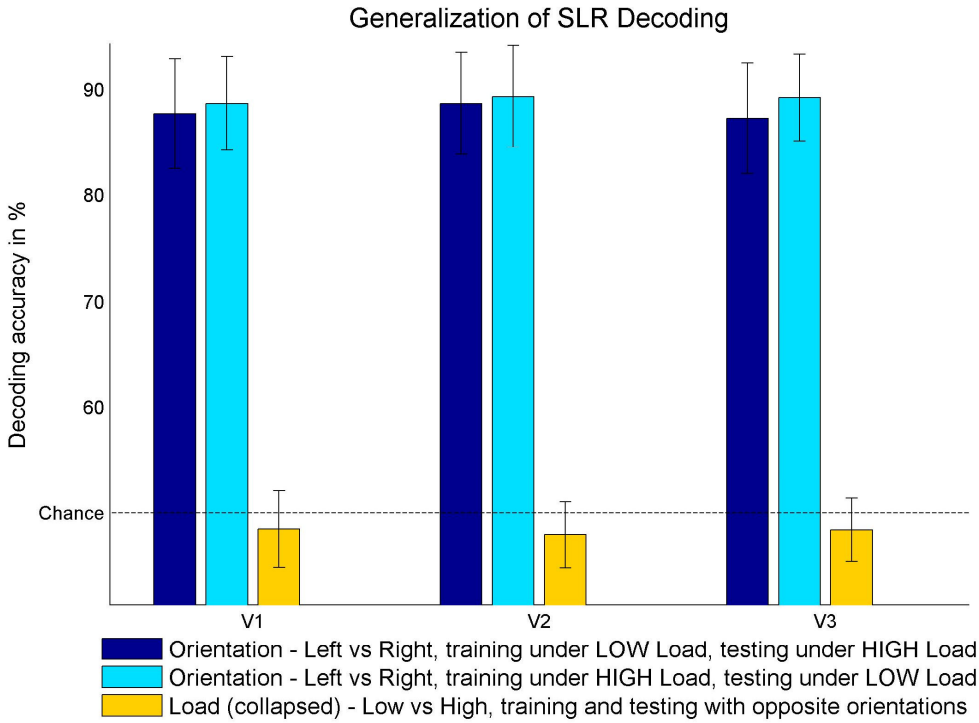



**Figure III-6**
**Generalization performance of SLR training. Shown in the figure are SLR decoding accuracies (mean across subjects) for visual areas V1 to V3 for test data different to the training data. Different colors indicate different combinations of generalization. Generalization is highly successful, compared to chance, under both load conditions (t test, all areas) with no significant differences between the two (paired t test, all areas). Generalization was unsuccessful for the attempt to generalize low vs. high decoding from different orientations (t test, all areas). Errorbars denote SEM across subjects.**

*Generalization Results*. Given training data of a pair of experimental conditions (i.e. left under low load vs. right under low load), I then tested how well this data set can be used to decode data from a different pair of experimental conditions (i.e. left vs. right, both under high load). This generalization decoding performance was computed using training on data from low load orientation data (left vs. right under low load) and testing on high load orientation data (left vs. right under high load), and vice versa. In both these out-of-load-generalization-decoding cases the resulting accuracy was significantly above chance [t-test, V1: training with low Load: 87.7%, $t(7) = 11$, $P < 0.01$; High Load: 88.7%, $t(7) = 9.8$, $P <$

0.01; V2: Low Load: 88.7%, t(7) = 9.5, P < 0.01; High Load: 89.4%, t(7) = 12, P < 0.01; V3: Low Load: 87.3%, t(7) = 8.8, P < 0.01; High Load: 89.3%, t(7) = 11, P < 0.01]. Overall, out-of-load-decoding results were very similar, and not significantly different, to results of decoding within-load [t-test, training low load, Testing: V1: Within load condition: 88.2%, Out of load condition: 89.4%, t(7) = 0.32, P = 0.75; V2: Within load condition: 86.6%, Out of load condition: 91%, t(7) = -0.97, P = 0.36; V3: Within load condition: 86.3%, Out of load condition: 90.9%, t(7) = -0.6, P = 0.56]. The same is true for training with high load data, testing on low load data [paired t-test, training high load: V1: Within load condition: 88.2%, Out of load condition: 89.4%, t(7) = 0.31, P = 0.76; V2: Within load condition: 86.6%, Out of load condition: 91%, t(7) = 0.89, P = 0.4; V3: Within load condition: 86.3%, Out of load condition: 90.9%, t(7) = 1.4, P = 0.2].

The same generalization approach applied to the prediction of low vs. high load (Training on leftwards orientations and testing on rightwards orientations, and vice versa) did not yield any results significantly different from chance [t-test collapsed across Orientations, V1: 48.5%, t(15) = 0.57, P = 0.58; V2: 47.9%, t(15) = -0.12, P = 0.9; V3: 48.4%, t(15) = 0.5, P = 0.6].

## III.4 Discussion

In the current study, I tested a number of hypotheses. Combining elements of experimental designs previously investigated (Haynes and Rees, 2005a; Schwartz et al., 2005), I was able to replicate a number of previous results, namely a behavioral manipulation with the central load task, a difference in distractor related brain activity across early visual cortices due to this load manipulation and successful decoding of peripheral oriented lines with MVPD. My third hypothesis was that there would be a reduction in MVPD-accuracy under high load, analogous to the reduced mean brain activity. However, orientation-classification under high load and low load was statistically indistinguishable. My fourth hypothesis was that there would be a significantly reduced ability to classify different load conditions instead of different orientations, and indeed results of load-prediction were below 55% accuracy. Yet, comparing high vs. low load classification still proved robust enough for a significant different from chance in V2 and V3. My final hypothesis proved entirely correct: it was possible to generalize from a training data in one load condition to test data taken from another load condition.

Retinal input was identical in both load conditions due to equal number of blocks of left and right tilted lines in the distractor annulus and a visually identical central task. Thus, differences between visual stimulation between experimental conditions were highly unlikely to account for any of my findings.

Behavioral findings confirmed that the attentional load manipulation in the central task was effective. It was much more difficult (attention demanding) to perform the task in the high-load condition than in the low-load condition for every participant. This replicates behavioural results and demonstrates a successful load manipulation (Schwartz et al., 2005).

A repeated measures ANOVA comparing brain activity evoked in early visual area V1 by peripheral orientations did not reveal any differences in brain activation comparing the activity evoked by task-irrelevant left tilts and right tilts (i.e. different visual features). However, the insignificant trend in V1 toward reduced overall brain activity elicited by the task-irrelevant peripheral distractor became significant in areas V2 and V3. This finding is consistent with previous findings using slightly different peripheral distractor stimuli (oriented lines vs. checkerboards) but the same central load task (Schwartz et al., 2005, Fig 7C). Like in the present study, Schwartz et al. (2005) show an increasing reduction in BOLD signal under the influence of load from V1 to V3 at a similar stimulus eccentricity. The same RM-ANOVA also revealed no significant differences between the different orientations. This is consistent with studies that report an inability to find feature-specific differences for oriented stimuli in visual cortex using conventional univariate analysis (Haynes and Rees, 2005a; Kamitani and Tong, 2005b; Haynes and Rees, 2006).

Multivariate analysis allowed accurate classification of the neural representation of one of two orthogonal orientations significantly

above chance. These results replicate recent results (Haynes and Rees, 2005a) with overall improved accuracy. Leftwards vs. rightwards distractor orientations under low and high load classify approximately equal in V1 to V3, with an accuracy of about 9 in 10 volumes classified correctly. Given the noisy nature of the BOLD signal, the lack of preprocessing steps compared to standard univariate analysis and previous results with only a minimal central task (Haynes and Rees, 2005a) the accuracy of decoding observed here is remarkable.

Since decoding was performed identically for high and low load conditions, it was feasible to compare the accuracy of different decoding results. However, comparing MVPD results from low and high load did not reveal any significant differences between the load conditions. This was surprising given the strong and successful behavioral manipulation of the central load task (Fig III-2) and the univariate reduction in overall BOLD signal (Fig III-4). Yet, it seems that differences in the task-irrelevant distractor representation are either not present or not measured by MVPD. Load theory predicts neural signals associated with task-irrelevant distractors (in our case the orientated lines) will be reduced when attentional resources are taken up by another task (in our case under high load). Thus, I hypothesized that classification accuracy revealed by MVPD should fall under high central load, especially having successfully replicated a behavioral modulation and a main effect of load for brain activity in early visual cortex. Yet, the results are not consistent with this hypothesis: high and low load decoding were indistinguishable. The overall accuracy was very high with

accuracies around 90%. Thus, it could be that classification in both cases reached a performance ceiling obscuring any small differences between the load conditions. The noisy nature of the BOLD signal may simply not allow any higher classification than about 90% explaining the null finding for the main effect of load for decoding. However, taking into account the strong behavioral modulation, the task being substantially more difficult under high load, this option seems unlikely.

MVPD was only marginally successful in decoding load from early visual cortex with accuracies only around 54% (Fig 5, yellow bars). Note that the one load-value reported is collapsed across results, however, all decoding-analysis was always performed on two conditions at a time (only results are collapsed for load, no collapsing happened during the actual analysis). Across all observations these relatively low accuracies were significantly different from chance in V1 and V2. Although I hypothesized that decoding accuracies should be significantly reduced or not significant at all, the result revealed statistically robust decoding of load in V1 and V2. This is particularly surprising as load decoding failed in V3, which means that these decoding results cannot be explained by a main effect in brain activity (which is significantly reduced in V3). However, disregarding the low accuracies, in case of V1 and V2, MVPD was sensitive enough to decode load significantly different from chance.

A number of possibilities could potentially account for the null-finding for a difference in decoding between different load conditions

and the somewhat surprising successful decoding of load in V1 and V2. It is important to understand the basis by which MVPD is deciding between different conditions, which is ultimately the voxel selection upon which any classification is based on. Here we used MVPD using ARD-SLR (automatic relevance detection for sparse logistic regression, for details, see Chapter II: General Methods). Studying the results for this experiment, it becomes clear that voxels selected relevant for classification are a poor representation of mean brain activity of the entire ROI. This is true for orientation classification (mean brain activity undistinguishable but decoding highly successful) and load classification (mean brain activity significantly different in V2 and V3 but decoding significantly different from chance in V1 and V2). Hence, the relevant voxels must reflect a different aspect of the data altogether. The location of a voxel decides which underlying neural activity it represents. For orientation, it has been suggested that an unequal distribution of orientation columns results in biased voxels distributed randomly across early visual cortex (Haynes and Rees, 2005a; Kamitani and Tong, 2005b; Haynes and Rees, 2006). However, a small number of invidividual MRI voxels might also capture a biased sample of orientation columns explained by a larger scale retinotopic radial bias rather than an anisotropic distribution of orientation columns alone (Sasaki et al., 2006). However, recent advances on single voxels corroborate and extend the claim towards an anisotropic distribution further by demonstrating voxel-based tuning functions to different orientations (Serences et al., 2009b). Consequently, it seems highly likely that voxels selected in our study (by ARD) also represent orientation biases for the two orientations we tested.

Thus, I conclude that, for our study, there are spatial areas in early visual cortex that, locally, show a clear and reproducible main effect for orientation that can be extracted by a spatial resolution of 3x3x3mm. These local differences are, however, not distinguishable when looking at the mean of all voxels. Consequently, voxels with increased activity must be counteracted by voxels with decreased activity. Comparing results from Serences et al (2009), orientation-biased voxels show roughly equal increase/ decrease of activity at 90 degrees (see Serences et al., 2009b, Figure 5). Therefore, voxels selected during orientation decoding are likely to be tuned voxels, increasing activity for their preferred orientation while decreasing activity for their anti-preferred direction.

Furthermore, in the case of load prediction, it seems logical that the strong main effect of mean brain activity in early visual cortex is only very weakly reflected in a spatial manner that our voxel selection grid (3x3x3mm) was able to extract. Potentially, load could still be spatially realized, but too densely, in early visual cortex. More likely, though, it is not at all a spatially localized effect. That would in turn mean that the load effect of mean activity is due to changes in overall activity rather than driven by single voxels, thus, voxels are not tuned to attentional load in an unrelated central task. This conclusion goes slightly against our surprising result of statistically robust load classification, which, therefore, remains unexplained. It is, nevertheless, only possible to classify load in fewer than 54% of all tested volumes.

As a final analysis, I successfully classified orientation from one load condition when training the MVPD algorithm with data from the other load condition and vice versa (Fig III-6). Results were just as high as during within load classification. This result is yet another indicator of the chosen voxels actually being biased by presented orientation, reliably across different conditions and independent to load in the central task. In contrast to this, voxels producing significantly above chance (V1 and V2) decoding for load conditions using one orientation only, did not generalize to load decoding with data from the other orientation (Fig III-6, yellow bars). This indicates that the chosen voxels in this case, most likely, were not biased to the decoded feature of low or high load. Again this further corroborates the argument that load seems not spatially distributed in early visual cortex. For a future analysis, it could be potentially beneficial to train the MVPD algorithm with data from different conditions that overlap in one feature that is tested for. This conjunction multivariate analysis could potentially help to identify biased voxels more clearly as interferences might even out and there could, potentially, be more training examples.

Overall, studies of perceptual load generally find decreased behavioral, perceptual and physiological measures of unrelated distractors (Lavie, 1995; Rees et al., 1997; Lavie and Fox, 2000; for review see Driver, 2001; for load theory see Lavie et al., 2004). However, here we found that orientation specific biases in brain activity of early visual cortex are independent of differential attentional load in a central task, even though the overall mean brain activity is decreased. Thus, as a result of the combination of

these two results and at least of the feature orientation, perception of a distractor that is decreased due to high attentional load in an unrelated task is not reduced due to the fact that basic features of the distractor are less clearly represented in early visual cortex. Any differences in perception must therefore be due to another factor as feature representation as measured by MVPD is unaffected by load in an unrelated task.

## III.5 Conclusion

The differences in MVPD and univariate analyses results highlight the complimentary nature of the two. Strengths of the MVPD method lie in distributed feature-specific analysis: i.e. different orientations are known to be represented in different neuronal sub-populations in early visual cortex (Hubel and Wiesel, 1968), resulting in an uneven spatial pattern that can be accessed successfully by MVPD. Strengths of univariate analysis methods lie in large scale, overall BOLD changes: i.e. load, unknown to have different underlying neuronal sub-populations, is well-known to reduce distractor activity in early visual cortices (Lavie, 2005; Schwartz et al., 2005) and led to a significant reduction of the BOLD signal in the present study.

Load theory predicts that the task irrelevant representation of the distractor (in this case orientations) should be significantly reduced when attentional resources are more bound to another task. Like previous fMRI experiments we identified a reduction in the

distractor related BOLD activity. However, we did not observe any significant differences in the underlying feature representation as measured by MVPD. Since this result is reported in conjunction with a qualitative replication of previous findings, it can help understanding about the underlying biological basis of the influence of attention on processing, especially feature specific, in the brain. From results of this study we conclude that any differences due to attentional load in an unrelated task of perception of orientation in a distractor are not due to the representation of orientation in early visual cortex as measured by MVPD.

A simple possibility to criticize any wider claims of results reported in this chapter is that the effect observed is purely restricted to the biological representation of orientation. Only a qualitative replication including another basic feature representation (e.g. colour or direction of motion) is able to dismiss this criticism. Therefore, the next chapter will explore if the observed result is restricted to orientation or can be extended to other features known to be represented in early visual cortex.

# Chapter IV:

# Effects of attentional load on motion selective processing I

## IV.1 Introduction

Classification accuracy for orientation of an unattended, irrelevant distractor remained unchanged under high (versus low) attentional load in a central task (chapter III). Thus, attentional load in a central task did not affect the representation of orientation of an irrelevant distractor. Yet, there is an ever growing body of studies showing that processing of irrelevant distractors is reduced under high attentional load. Such experiments span an incredibly wide variety from the classical effects on distractors consisting of color-shape conjunctions (for example: Lavie, 1997), to measures of BOLD signals in visual cortex of basic visual features (Rees et al., 1997), or face specific BOLD and even working memory related measures (Jenkins *et al.*, 2003, 2005). More recently high load was additionally shown to reduce effects of individual differences in distractibility (Forster and Lavie, 2007) and the effect of high load was shown to also extend to relevant distractors (Forster and Lavie, 2008). Taking into account this vast body of studies the result of chapter III was not predicted by load theory. But to further understand this unexpected result, it is important to understand if the result observed in chapter III is specific to the stimulus used, to

the analysis performed or if it could indeed be generalized to other types of visual features and their neural representations.

### IV.1.1 Objectives

Here, I present an experimental paradigm designed to closely match the one presented in chapter III. However, instead of measuring BOLD signal associated with the basic visual feature of orientation, I investigated the effect of load to the basic visual feature of direction of motion. This was achieved by replacing the oriented distractor stimuli with a field of dots moving in one of two possible directions. Thus, the visual feature orientation is replaced with the feature motion direction which, similar to orientation, has been shown previously to be distinguishable with multivariate pattern recognition applied to BOLD contrast signals from visual cortex (Kamitani and Tong, 2005a).

For direction of motion, neurons in the middle-temporal area (V5/MT) of monkeys are selectively responsive to the direction of motion (Zeki, 1983; Albright et al., 1984; Zeki et al., 1991). Further research established that monkey area V5/MT represents different directions of motion in a columnar architecture (Tootell et al., 1995; Tootell and Taylor, 1995) and there is evidence that this might also be true for humans, mainly from fMRI-adaptation studies (Heeger et al., 1999; Nishida et al., 2003; Seiffert et al., 2003). Therefore, just like orientation and the early visual cortex, area V5/MT (alongside

V1 to V3) was an obvious area to study the effect of varying attentional load.

### IV.1.2 Hypothesis

The experiment presented here sought firstly to successfully decode the direction of motion of unattended random moving dots.

Secondly, I sought to replicate the behavioral load effect of the central task again just as seen in Chapter III.

Further, for the difference between attentional load conditions, my goal for the experiment presented here was to identify any effects of central attentional load brain activity associated with irrelevant (distractor) motion. Thus, my third hypothesis was to find a differential representation of feature specific signal (motion direction left-upwards vs. right-upwards) in neural activity of early visual cortices V1-V3 and area V5/MT under high load (compared to low load). Note that, due to the high similarities to the experiment presented in chapter III, failure to observe such central-load-depended differences in results from multivariate pattern decoding (MVPD) of (motion direction specific) distractor activity would, however, be less surprising than before.

Finally, my last hypothesis was that it would be possible to predict load (instead of motion direction) from early visual cortices, analogous to the same experimental question in chapter III.

## IV.2 Methods

### IV.2.1 Participants

Eight healthy participants (6 male, mean age 28 years) gave written informed consent to participate in the study, which was approved by the local ethics committee. All participants had normal or corrected to normal vision. One participant had to be excluded from the analysis due to her paying no attention during the central load task and consequently failing to comply with the task requirements, another participant was excluded due to excessive head movements during scanning.

### IV.2.2 Stimulus

Participants performed a visual detection task similar to that described in chapter III. In brief, a continuous rapid successive visual presentation (RSVP) of colored crosses (one cross every 750 ms) was shown at fixation in the centre of the screen (Figure IV-1). This RSVP stream consisted of cross-shaped stimuli with two different orientations (upright or upside-down) and six different colors in random order (Figure IV-1 B). Participants were required to monitor the occurrence of infrequent (14%) pre-specified targets within this rapid central letter stream, and to respond by a button-press to each detected target. Reaction times (RT) were collected as behavioral responses with a standard MRI-button box. Different to the central task in Chapter III the length of every experimental

block within each run was decreased to 17 seconds. Accordingly, the frequency of targets was slightly increased to allow a minimum of two targets per block. Both changes were tested behaviorally before scanning and did not change the efficiency of the load task. The low-load (color) task required a key-press for any red cross irrespective of its orientation, whereas the high-load (conjunction) task required a key-press for any upright yellow cross or upside-down green cross. Items that were targets in one condition also appeared with the same frequency as task-irrelevant stimuli in the other condition (i.e. high-load targets appeared as distractors under low-load instructions, or vice versa). Thus, only task instructions distinguished the high-load and low-load conditions for the central task.

Before the first and after each experimental block, the RSVP task was replaced by a small fixation spot on the same medium gray background used during the experiment. These rest blocks were 3 seconds long before the first block and lasted 17 seconds between experimental blocks.

During experimental blocks, in addition to the cross task, the visual stimuli comprised an annulus shaped field with a radius between 3° and 10°. This field contained 100% coherently moving black dots (for maximum contrast) on a medium gray background (Figure IV-1 A). During the experimental conditions there were always exactly 1400 dots visible, moving either 45° tilted or -45° tilted upwards (further referred to as left-upwards and right-upwards movement). Dot lifetime was limited to 350ms to ensure that it was easy for

participants to keep fixation on the central task. Each dot moved at a constant velocity of 5° per second. At the end of its lifetime or when a dot reached the border of the stimulus it was immediately replaced by a new dot in a random location. The field of moving dots is further referred to as irrelevant distractor since it was completely irrelevant to the central task and participants were instructed only to attend to the central task.

### IV.2.3 Procedure

Participants lay supine in the scanner and viewed visual stimuli that were projected from an LCD projector (NEC LT158, refresh rate 60 Hz) onto a screen viewed via a mirror positioned within the MR head coil. Stimuli were presented using MATLAB (Mathworks Inc.) and COGENT 2000 (www.vislab.ucl.ac.uk/Cogent/index.html). Complete darkness was achieved in the scanning environment by manually masking the fMRI projection screen, head coil and internal bore with matt black card. This eliminated discernable non-retinotopic luminance cues, ensuring that the only source of visual stimulation during experimental runs was the experimental stimulus.

All participants were scanned for a total of 8 (n = 2) or 10 experimental scans (n = 6). Each scan consisted of four experimental blocks alternating every combination of motion direction and difficulty of the central task in a manner that was counterbalanced across runs and participants. Compared with the stimuli described in experimental chapter one, the timing of each

block was slightly altered. Each scan started with the appearance of the central fixation spot for 7s including a short reminder of the task to come, followed by a 17s long block of the load task at central fixation with the motion stimulus (irrelevant distractor) surrounding the central task. Then 17s of rest were used to give participants a short rest before the next experimental block. Participants were instructed to concentrate on the central task/ fixation cross for the duration of the entire run.

Each experimental run lasted 104 volumes resulting into 136 seconds, split into 4 blocks of 34s seconds each (17 seconds task, 17 seconds rest). Hence, each experimental fMRI run comprised each of the four task conditions once (low/ high load combined with left-upwards and right-upwards movement).

### IV.2.4 Scanning

*Imaging Parameters.* A 3T Siemens Allegra system acquired T2*-weighted Blood Oxygenation Level Dependent (BOLD) contrast image volumes using a descending sequence every 1.3s. Each volume comprised 20 slices with a slice thickness of 3mm, positioned on a per participant basis to give coverage of the occipital lobe with an in-plane resolution of 3x3 mm. To maximize signal to noise in early visual cortex an occipital head coil was used.

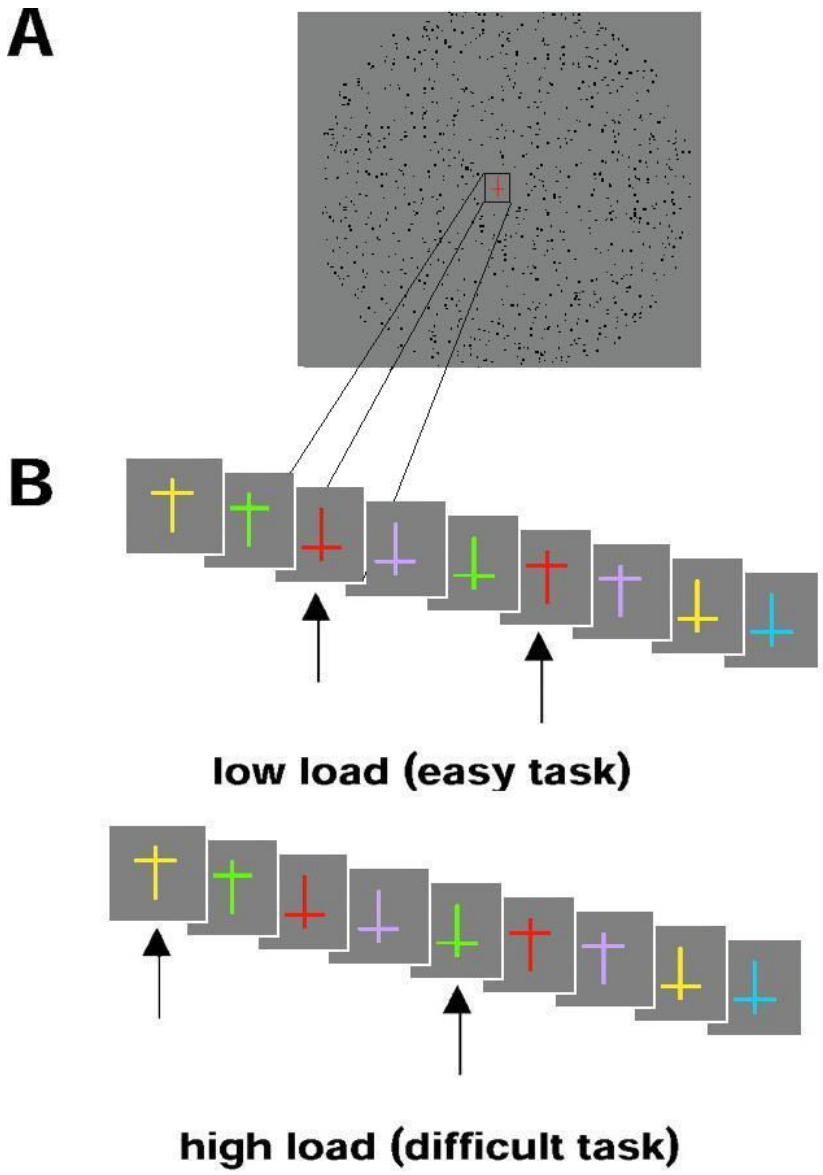



**Figure IV-1**
**Example stimuli used in the experiment**
**A) Stimuli used in the main experiment consisted of a random dot field moving either left-upwards or right-upwards at an angle of 45 degree. At fixation an attentional load task, identical to the load task described in chapter III, is performed by the subject.**

**B) The load task RSVP as performed in the main experiment. Subjects were required to detect red crosses either upright or inverted (easy task, low attentional load) or the yellow upright and an inverted green crosses (difficult, conjunction task, high attentional load). See Methods for further details.**

*Identification of ROIs.* To identify the boundaries of primary visual cortex (V1) and extra-striate retinotopic cortical areas V2 and V3 on a participant by participant basis, standard retinotopic mapping stimuli were presented twice to each participant (Sereno *et al.*, 1995). For V1 to V3, I followed standard segmentation and cortical flattening in MrGray (Teo *et al.*, 1997; Wandell *et al.*, 2000) to determine the borders between the ROIs on a flattened cortical representation (see Chapter III for details). Finally, to identify motion sensitive area V5/MT, an MT-localizer sequence was applied twice per participant. During these scans participants solely had to fixate on a central fixation cross while, alternating, 15 volumes long intervals of expanding and contracting moving dots and 15 volumes long intervals of static dots were displayed. These moving dots were eccentricity matched to the conditions of moving dots during the main experiment. Individually thresholded contrasts of moving vs. static dots formed the basis for the definition of an 8mm diameter sphere coving peak activity. The overlap of this sphere and gray matter, as determined by SPM5 segmentation, formed the basis for V5/MT definition. Both, retinotopic mapping runs and V5/MT localizer runs lasted for 205 image volumes.

In total, 12 to 14 scanning runs of 104 to 205 image volumes were acquired per participant.

### IV.2.5 Analysis

*Data preprocessing.* Data were preprocessed using Statistical Parametric Mapping software (SPM5, www.fil.ion.ucl.ac.uk/spm5). After discarding the first three image volumes from each scan to allow for T1 equilibration effects, functional image volumes were realigned to the mean of the remaining volumes. Following that the structural of each participant was and co-registered with the individual participants' realigned data. All data was also high-pass filtered (cut-off 128s) to remove low-frequency signal drifts. For multivariate analysis experimental data was not further preprocessed. However, for further univariate analysis with SPM5 a copy of the data was spatially smoothed with a kernel of full width half maximum of 5mm and normalized to standard Talairach space (Tournoux, 1988).

*Univariate analysis with SPM5.* Data from the main experiment was further analyzed using a general linear model (GLM). We used a GLM containing boxcar waveforms representing each of our experimental conditions, convolved with a canonical hemodynamic response function (HRF). In all contrasts reported, low and high load contain equal numbers of trials from experimental conditions. First level analysis was performed setting up T-contrasts between the different conditions (regressors) for each participant individually. First level contrasts were set up for the spatially normalized data and then further analyzed on the group level with a second level analysis step, setting up a t-test for individual contrasts between all participants.

With the first, realigned only, data set I proceeded identically to Chapter III: Data from 8-10 scans of 4 blocks for each participant was first averaged over all voxels in a ROI, in all volumes of a block. Then I averaged these block-values across scans, resulting in 4 values of mean brain activity per participant. Finally, I compared different conditions of brain activity, i.e. high vs. low load collapsing across orientation, with paired t-tests.

*Multivariate pattern classification.* Pattern classification was performed using Sparse Logistic Regression (SLR, see Chapter II: General Methods). For determining classification accuracy, only classification with unseen and independent-test data was considered. Thus, test data was always independent from the training data. The actual classification was repeated n times, where n was determined by the number of independent blocks (n-fold-cross-validation). Classification accuracies were then averaged across these n data assignments.

Similar to previous applications of MVPD (Kamitani and Tong, 2006), multivariate pattern classification was used here, to predict the direction of motion displayed in the distractor annulus. After predicting direction of motion, MVPD was repeated, this time round attempting to classify load (similar to Chapter III). After repeating MVPD for each participant, I tested for an overall significant difference from chance (50%) with a simple student's t-test.

Different from methods described in Chapter III, I introduced two crucial new elements into the analysis. Firstly, inspired by the generalization result for orientation decoding (Chapter III, Figure III-6), I pooled training data from both load conditions, instead of keeping training data for low and high load strictly independent. Thus, I trained the SLR algorithm with double the number of upwards-left versus upwards-right examples, irrespective of the load condition of the central task. During testing, however, I kept-testing data strictly separate for low and high load, resulting in two different decoding results achieved upon the same training data. To test for the validity of this procedure I also produced the results with independent training & test data (identical to Chapter III).

Secondly, many recent publications report multivariate results based on averaged vectors across multiple volumes (Kamitani and Tong, 2005b, 2005a, 2006; Serences and Boynton, 2007a; Serences et al., 2009b). Others report classification results only on single volumes (Haynes and Rees, 2005b, 2005a, 2006, Chapter III). To compare these two approaches I first ran the entire MVPD analysis with single volumes (also called single TRs, Fig. IV-6) and then repeated the analysis with time-averaged volumes from entire blocks of 15 volumes (Block averages, Fig. IV-7).

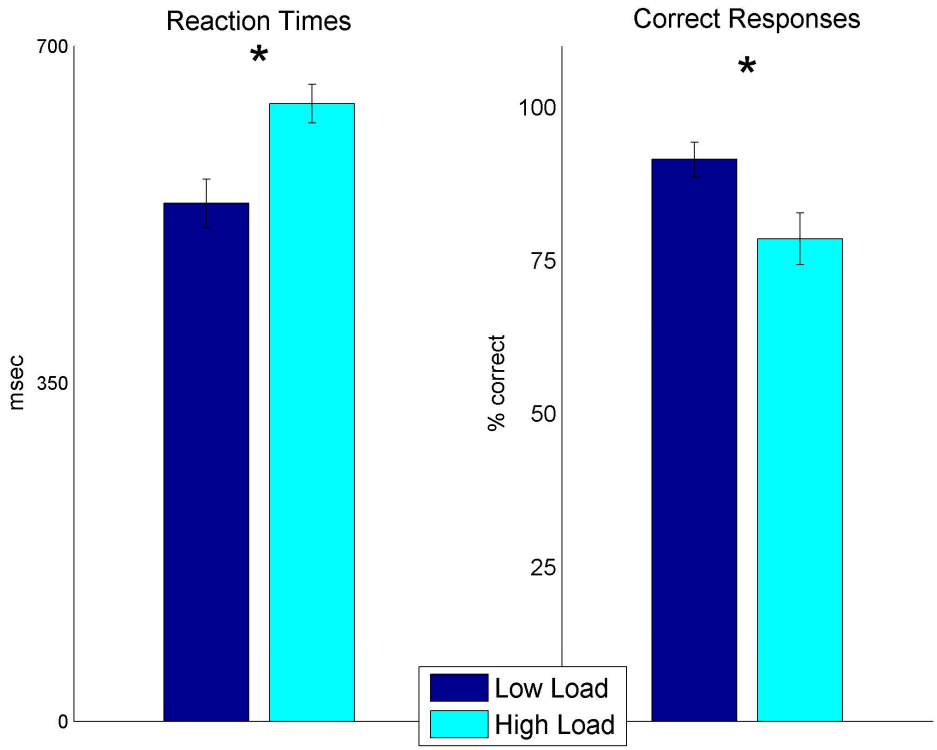



**Figure IV-2**
**Shown are behavioral results of performance during the load task. Mean reaction times of subjects for successful detection of target crosses are shown on the left. Mean correct hits across subjects for target crosses are displayed on the right. Separate colors denote low (dark blue) and high (light blue) load conditions in the central task. Both differences are highly significant (P< 0.001) and present in every subject. Errorbars indicate standard error of the mean (SEM) across scans (8-10 per subject).**

## IV.3 Results

### IV.3.1 Behavioral results

In both load conditions, performance was always above chance, confirming adherence to the task requirements. RSVP together with the unpredictability of targets ensured that participants always

monitored items at central fixation during both task conditions. I examined Reaction Times (RT) across participants comparing the RT of low load conditions with the RT of high-load conditions with a paired, two-tailed t-test. Mean detection latencies for central target crosses were significantly slower in the high-load versus low-load condition [mean reaction time high load high: 640ms; low load: 536ms; $t(46) = 7.9$, $P < 0.001$] (Figure IV-2, right).

I also tested hit rates of correctly detected crosses in both conditions. Again I discovered a significant difference between high and low load condition. Correct responses significantly decreased in the high load condition [mean correct responses high load: 78.5%; low load: 91.4%; $t(46) = -3.4$, $P < 0.001$] (figure IV-2, left).

Finally, I examined the error rates (incorrect target responses or false alarms, not including missed responses) using a similar paired t-test. I found a significant difference between the two conditions. False alarms significantly increased during the high load task [M ER high: 35.4%; low: 18.7%; $t(46) = 4.5$, $P < 0.001$].

Together, these findings confirm that attentional load was effectively manipulated by the central task.

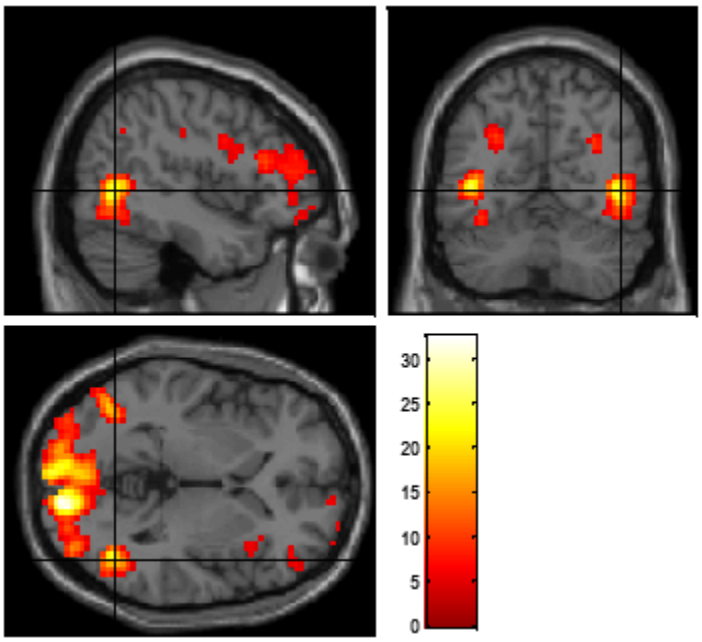



**Figure IV-3**
**Exemplar single subject functional data normalized and superimposed on a standard, normalized SPM structural. A random effects comparison at T= 4.73, p < 0.05 (Family wise error (FWE) corrected, no voxel threshold) reveals areas activated by the comparison of all blocks containing motion vs. rest blocks. Bilaterally, motion responsive V5/MT is visible (right V5/MT indicated with crosshair). Additionally, early visual cortices around the occipital pole show noticeable activity, demonstrating higher activity during experimental (task + motion) blocks than during rest blocks.**

### IV.3.2 Univariate Results

I examined the contrast of all blocks containing (irrelevant) motion vs. rest blocks. This contrast produced very clear activation in visual cortices and V5/MT on an individual participant basis at T = 4.73, p < 0.05 (Family wise error (FEW) corrected, exemplar participant Figure III-3). This pattern roughly repeats for each participant, however visual inspection of a random effects analysis on the group level (n=6), at a conservative T = 5.21, p <0.001 (uncorrected), shows no voxels at all above this threshold. At a more liberal

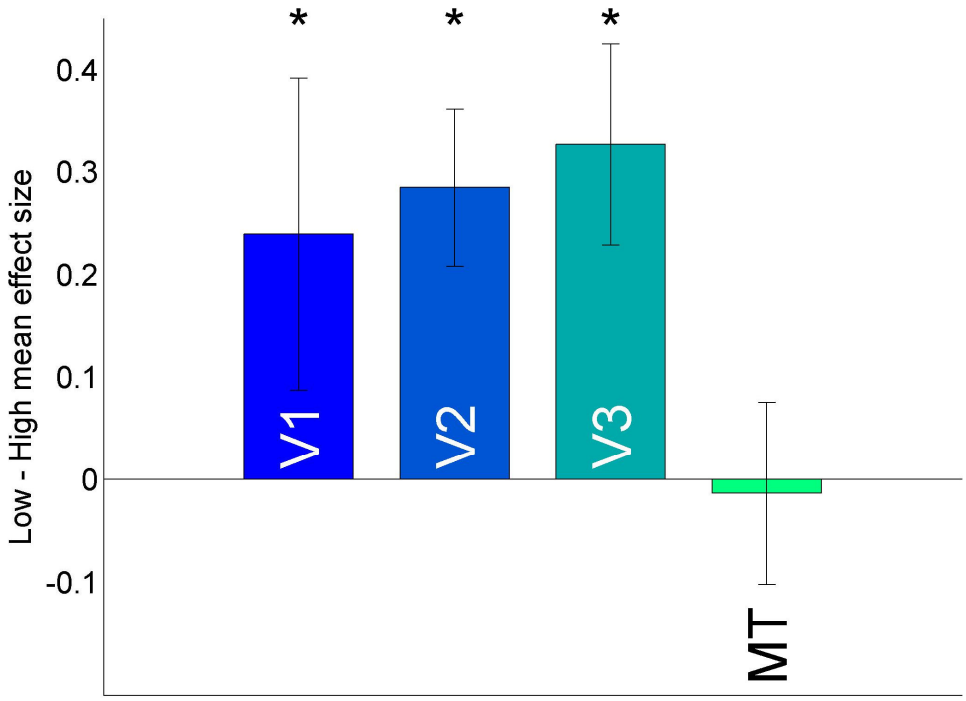



**Figure IV-4**
**The graph is showing the difference in mean activity in each area comparing high and low load, averaged across scans and subjects. Values are the mean brain activity of high load blocks subtracted from the mean low load activity. While activity is significantly reduced under high load for V1 to V3, there is no activity difference in V5/MT. Errorbars indicate standard error of the mean (SEM).**

threshold (T = 3.144, p <0.01 (uncorrected)), activity in visual cortices V1 to V3 and V5/MT bilaterally becomes clearly visible, echoing the single participant results on the group level. Figure IV-3 also shows a small activity patch around the right central sulcus which may reflect button pressing with the left hand during the experimental conditions (vs. no button pressing during rest).

I continued examining the result of the contrast low load motion vs. high load motion, in other words the main effect of low load (vs. high load) in the context of motion. This was done for each ROI, V1 to V3 and V5/MT, separately. All ROIs were pre-defined according to

activity as measured by independent localizer scans (see Methods for details). Ignoring the direction of the distractor motion, I calculated the value of mean brain activity for low load experimental blocks and high load experimental blocks for each participant. Thus for each participant I had a pair of brain activity values reflecting mean brain activity of low load and high load blocks. With a null-hypothesis of no differences between these two conditions, a paired t-test revealed significant differences between low and high load pairs. In V1 to V3 motion-related activity from low load blocks is significantly higher than activity in high load blocks. However, no such activity difference could be found in V5/MT [V1: Low load: 0.59, High load: 0.35, $t(11) = 3.0$, $P = 0.01$; V2: Low load: 0.63, High load: 0.34, $t(11) = 4.3$, $P < 0.01$; V3: Low load: 0.7, High load: 0.37, $t(11) = 4.9$, $P < 0.01$; MT: Low load: 0.95, High load: 0.96, $t(11) = -0.18$, $P = 0.86$]. Figure IV-4, illustrates these findings in form of a difference plot between BOLD signal from low load blocks minus activity from high load blocks. To double check this surprising lack of load related activity differences in V5/MT, I visually inspected a random effects analysis on the group level at an extremely liberal threshold of $T = 1.44$, $p < 0.1$ (uncorrected)). However, even then, V5/MT bilaterally showed no differences of low vs. high load motion, while early visual cortices clearly showed differential activity.

Thus, the main task (motion blocks) compared to rest blocks nicely activated brain areas related to the perception of motion (V1-V3 and V5/MT, Figure IV-3). In areas V1 to V3 there was additionally a motion-related activity difference between low and high load blocks

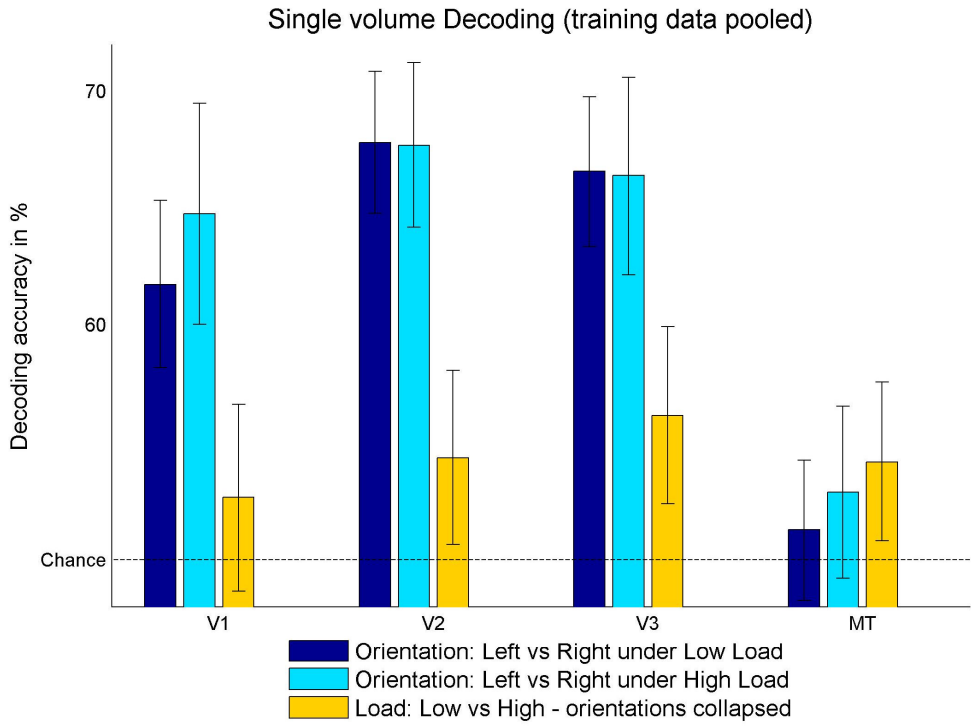



**Figure IV-5**
**Decoding accuracy (mean across subjects) for visual areas V1 to V3 and MT, obtained with Sparse Logistic Regression (SLR). Different colors indicate separate decoding performance for different test-comparisons: leftward vs. rightward orientation decoding was highly successful, compared to chance, under both low load (dark blue bars) and high load (light blue bars) conditions. In both conditions, training data (left vs. right) was pooled across load conditions to increase power of training. Paired t-tests between the two orientation decoding results revealed no significant differences in any area.**
**Load decoding (yellow bars), decoding low vs. high load rather than different directions of motion, proved significantly different from chance in areas V2, V3 and MT. Similarly to direction of motion results, training data was pooled, this time across motion directions, to increase training power. The reported result is the mean of separate testing for both (high vs. low with either left or right upwards moving dots). Errorbars denote mean of subject-specific SEMs.**

with activity decreasing under high load. However V5/MT does not show this difference in motion related activity (Figure IV-4). Thus, the same irrelevant motion perceived under either low load or high load produced no activity difference in bilateral V5/MT, but it does in V1 to V3.

### IV.3.3 Multivariate classification result

In this chapter, I report classification results obtained with the same SLR- MVPD analysis strategy as presented in Chapter III. Figure IV-5, Figure IV-6 and Figure IV-7 show decoding accuracies for areas V1 to V3 and V5/MT. The y-axis depicts mean accuracy of classification across participants, while different colors along the x-axis indicate decoding from different experimental conditions in the different visual areas. In Figure IV-5 and Figure IV-6 the training data was pooled to increase training power (see Chapter III). Thus, results obtained for low and high load are obtained from exactly the same voxel selection, only tested with volumes from either condition separately. To be sure this procedure was not falsifying results Figure IV-7 contains the same decoding outcome obtained without pooling of training data. Finally, for results of Figure IV-6 and Figure IV-7B, the input into the SLR algorithm was block averaged volumes rather than single volumes (see Methods).

In Figure IV-5 training and testing was performed on single volumes. Under both high load and low load, decoding left-upwards versus right-upwards direction of motion was successful (significantly different from chance) in V1 to V3. In V5/MT prediction accuracy was markedly reduced and close to chance. Decoding motion direction under low load was unsuccessful and only just significantly different from chance under high load [ V1: Low Load: 61.8%, $t(5) = 2.7$, $P = 0.04$; High Load: 64.8%, $t(5) = 5.3$, $P < 0.01$; V2: Low Load: 67.8%, $t(5) = 4.4$, $P < 0.01$; High Load: 67.7%, $t(5) = 5.8$, $P < 0.01$; V3: Low Load: 66.6%, $t(5) = 5.5$, $P <$

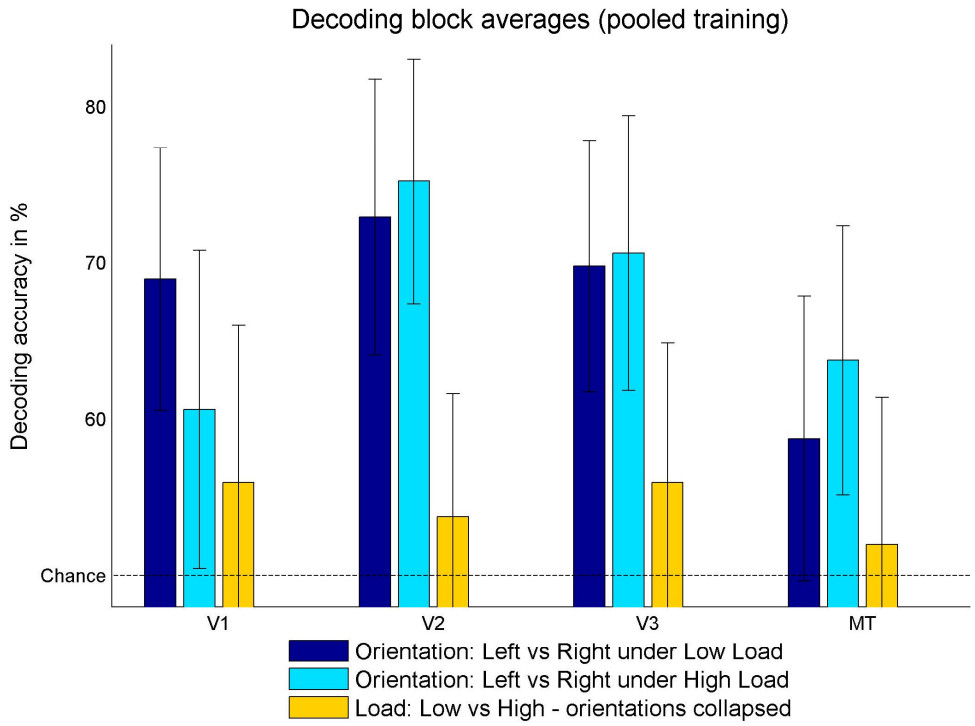



**Figure IV-6**
**Decoding block averages (mean across subjects) for visual areas V1 to V3 and V5/MT, identical to Figure IV-6 but with block averages instead of single volumes. Errorbars (mean of subject-specific SEMs) remarkably increase demonstrating a far higher volatility in the results (outlier results 33% of subjects). Yet, nearly all motion direction decoding accuracies increase due to block averaging (5,1% increase on average, one exception: V1, high load). In contrast to that, load decoding (yellow bars) accuracies decrease leaving load decoding statistically indistinguishable from chance in all ROIs.**

0.01; High Load: 66.4%, t(5) = 6.7, P < 0.01; MT: Low Load: 51.3%, t(5) = 0.54, P = 0.6; High Load: 52.9%, t(5) = 3.2, P = 0.03]. Decoding load proved successful (significantly different from chance) in V2, V3 and MT [Load decoding (collapsed): V1: 52.6%, t(11) = 2.1, P = 0.06; V2: 54.4%, t(11) = 3.6, P < 0.01; V3: 56.1%, t(11) = 4.5, P < 0.01; MT: 54.2%, t(11) = 3.3, P < 0.01]. Finally I compared low and high load decoding results for a difference in decoding performance. However, no visual area showed any significant difference between the two load conditions [ V1: Low Load 61.8%, High Load 64.8%, t(5) = -0.68, P = 0.52; V2:

Low Load 67.8%, High Load 67.7%, t(5) = 0.039, P = 0.97; V3: Low Load 66.6%, High Load 66.4%, t(5) = 0.059, P = 0.96; MT: Low Load 51.3%, High Load 52.9%, t(5) = -0.62, P = 0.56].

In Figure IV-7 training and testing was performed on block averages. This means each experimental block of 19.5 seconds (15 volumes) was averaged to one block-average-volume. Applying this technique, overall decoding accuracy increased by an average of 5.1% for motion direction decoding. However, together with the overall accuracy of decoding the volatility in the results also increased (see increase errorbars in Figure IV-7). Motion direction decoding was successful in high and low load in V1 to V3, unsuccessful in MT (likely to be due to increased SEM). Load decoding was unsuccessful in all ROIs [V1: Low Load: 69%, t(5) = 5.4, P < 0.01, High Load: 60.6%, t(5) = 2.7, P < 0.05; Orientation (collapsed): 55.9%, t(11) = 2, P = 0.07; V2: Low Load: 72.9%, t(5) = 3.7, P < 0.05; High Load: 75.2%, t(5) = 3.2, P < 0.05; Orientation (collapsed): 53.8%, t(11) = 1.9, P = 0.09; V3: Low Load: 69.8%, t(5) = 2.8, P < 0.05; High Load: 70.6%, t(5) = 3.8, P < 0.05; Orientation (collapsed): 55.9%, t(11) = 1.6, P = 0.15; MT: Low Load: 58.8%, t(5) = 1.3, P = 0.25; High Load: 63.8%, t(5) = 2.4, P = 0.06; Load (collapsed): 52%, t(11) = 0.53, P = 0.61]. Differences between low and high load motion decoding were not significant [V1: Low Load 69%, High Load 60.6%, t(5) = 1.6, P = 0.17469; V2: Low Load 72.9%, High Load 75.2%, t(5) = -0.4, P = 0.70711; V3: Low Load 69.8%, High Load 70.6%, t(5) = -0.13, P = 0.89909; MT: Low Load 58.8%, High Load 63.8%, t(5) = -0.51, P = 0.63271].

To ensure that pooling of training data was a valid methodological step and did not falsify results or decrease differences between high and low load, I repeated the entire analysis with completely separate training and testing for high and low load. Visually inspection of the results (Figure IV-8) reveals that all results have decreased decoding accuracy (about 6-10% lower), but also that the overall results largely remained the same. Single volume decoding (Figure IV-8 A) was still successful in V1 to V3 for high and low load, with the exception of low load classification in V1 which was also lowest in the corresponding result in Figure IV-6. Decoding was unsuccessful in MT [V1: Low Load: 52.1%, $t(5) = 0.73$, $P = 0.49$; High Load: 58.1%, $t(5) = 4.7$, $P < 0.01$; V2: Low Load: 57.9%, $t(5) = 2.5$, $P < 0.05$; High Load: 59.9%, $t(5) = 2.7$, $P = 0.04$; V3: Low Load: 59.6%, $t(5) = 3.5$, $P < 0.01$; High Load: 61.1%, $t(5) = 3.5$, $P < 0.05$; MT: Low Load: 53.8%, $t(5) = 1.6$, $P = 0.17$; High Load: 50.4%, $t(5) = 0.28$, $P = 0.79$]. There was no significant difference between high and low load [V1: Low Load 52.1%, High Load 58.1%, $t(5) = -2.3$, $P = 0.07$; V2: Low Load 57.9%, High Load 59.9%, $t(5) = -0.77$, $P = 0.47$; V3: Low Load 59.6%, High Load 61.1%, $t(5) = -0.49$, $P = 0.64$; MT: Low Load 53.8%, High Load 50.4%, $t(5) = 2.4$, $P = 0.059$].

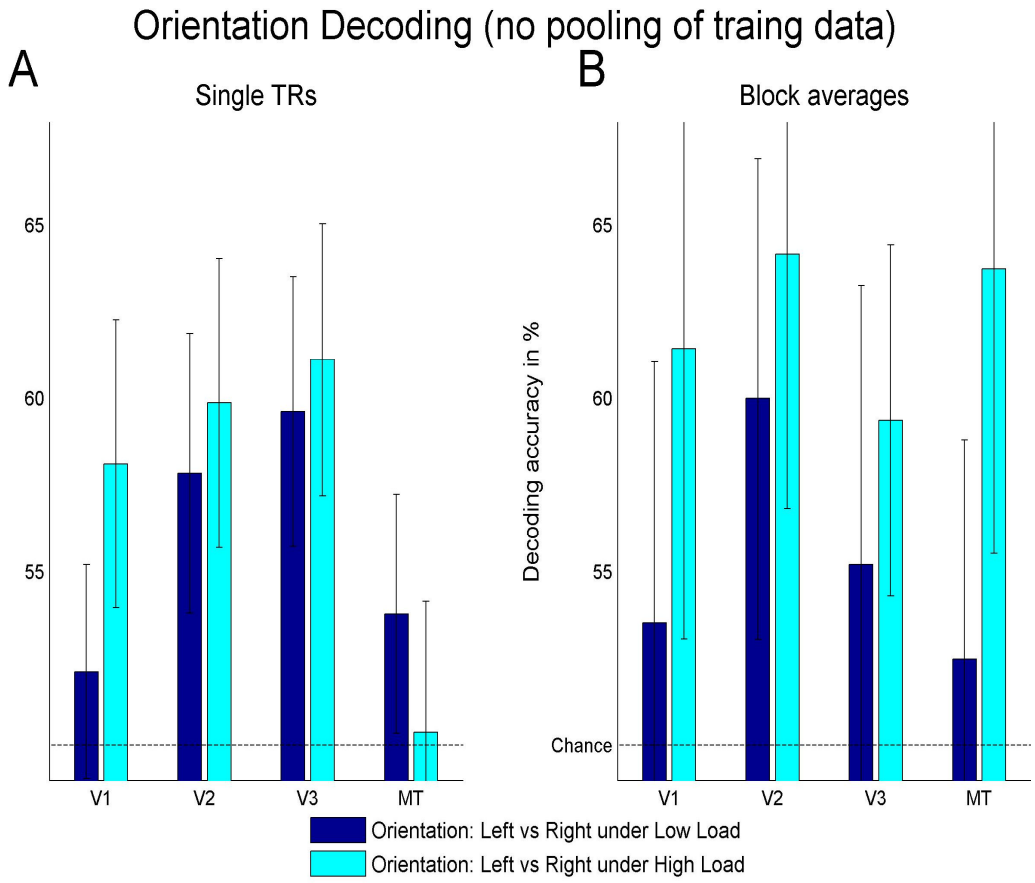



**Figure IV-7**
**SLR Decoding accuracy (mean across subjects) for visual areas V1 to V3 and MT. Here, training for decoding was only performed with data from one load condition at a time (separate training & testing).**
**Different colors indicate separate decoding performance for leftward vs. rightward orientation decoding in low (dark blue) or high load (light blue). Errorbars denote mean of subject-specific SEMs. Overall, results confirm trends observed from results with pooled training (Figure IV-6,IV-7), albeit with decreased decoding performances (about 6-10% lower), likely owing to less accurate training due to lack of power (only half as many training volumes). For details see Discussion.**
**A: Training and testing on single volumes (TRs), analog to Figure IV-6. Even though accuracies are reduced, decoding was still successfully different from chance in high load (V1-V3) and low load (only V3). Differences between low and high load decoding remained non-significant.**
**B: Training and Testing on block averages, analog to Figure IV-7. Due to the increased volatility (note large errorbars) and the decreased overall performance, no result was significantly different from chance anymore.**

For block average decoding the data appeared visually comparable to training with pooled data (Figure IV-7), however due to the

increased SEM no result is significantly different from chance anymore but decoding high load decoding in MT [V1: Low Load: 53.5%, t(5) = 0.72, P = 0.50; High Load: 61.5%, t(5) = 1.4, P = 0.217; V2: Low Load: 60%, t(5) = 0.88, P = 0.42; High Load: 64.2%, t(5) = 1.9, P = 0.12; V3: Low Load: 55.2%, t(5) = 1.1, P = 0.31; High Load: 59.4%, t(5) = 1.3, P = 0.26; MT: Low Load: 52.5%, t(5) = 0.49, P = 0.65; High Load: 63.8%, t(5) = 3.5, P = 0.02]. All differences between low and high load are not significant [V1: Low Load 53.5%, High Load 61.5%, t(5) = -1, P = 0.36; V2: Low Load 60%, High Load 64.2%, t(5) = -0.46, P = 0.66 V3: Low Load 55.2%, High Load 59.4%, t(5) = -0.58, P = 0.59; MT: Low Load 52.5%, High Load 63.8%, t(5) = -1.8, P = 0.14].

## IV.4 Discussion

*Behavioral results*. Even though it was slightly altered from Chapter III, the central task was attentionally more demanding for the high-load than low-load condition for all participants. The behavioral results confirmed that central attentional load was effectively varied by our task manipulation. Furthermore, the similarity of the behavioral measures to the findings obtained in previous applications of the task (Schwartz et al., 2005, chapter III) strongly suggested that the slight alterations to the task did not change the effectiveness of the load manipulation.

*Load effects on fMRI brain activity.* The simple contrast of all motion vs. baseline nicely confirmed previous results of activity in early

visual cortices and in V5/MT (Sereno et al., 1995; Tootell et al., 1995; Tootell and Taylor, 1995; Tootell et al., 1996) (Figure IV-3). However, more interestingly, motion-distractor related activity was decreased under high load in early visual cortices V1 to V3. However, for the motion sensitive area V5/MT this was not true (Figure IV-4). Instead, there was no difference for activity evoked by the task irrelevant distractor motion under the high and low load in V5/MT. Even though there were many differences to experimental paradigms of other studies that studied distractor activity under differing load conditions (O'Connor et al., 2002; Jenkins et al., 2003; Lavie, 2005; Schwartz et al., 2005), this apparent lack of decreased activity under high load in V5/MT was surprising. One previous study specifically studied the representation of task irrelevant motion under load (Rees et al., 1997) and found a decrease of activity in V5/MT for motion (vs. no motion) under low load (vs. high load). However, a close comparison between the present and previous studies revealed the crucial difference that the experimental design reported here lacked a condition with the central task under high and low load and static dots instead of moving dots - a 'task-only, no-motion' condition. It was therefore impossible to replicate the exact interaction effect reported in Rees et al. (1997). Consequently, this, to some extent unexpected, result could be fully explained with an additional experiment including such additional no-motion conditions. Chapter V will explore such an experiment. For this study, however, without a task-only, no-motion condition, we found there were no differences in V5/MT activity comparing motion related activity under high and low load.

Another possibility not easily dismissed is that motion, as presented in our stimuli, might have elicited more eye movements in one of the two load conditions. These eye-movements, if they occurred, might explain some of the effects observed. Motion is known to attract tracking eye movements and, therefore, eye movements offer an easy general criticism. Behavioral results did not indicate eye movements and, in fact, the rapid successive visual presentation task required constant monitoring of the crosses and, thus, does not allow for eye movements. Yet, only a replication of the experiment that includes eye tracking could truly eradicate this possibility. Again, chapter V explores this possible criticism by measuring eye-movements.

*Methodological MVPD aspects.* Taking the spatial pattern of activity in visual cortex into consideration, multivariate analysis allowed accurate classification of the neural representation of one of two motion directions. This was true despite the fact classification was performed on an ignored distractor containing the decoded feature. Classification was successful in early visual cortices V1-V3 with single volumes and with block averages, in V5/MT only with block averages. Inspired by the successful generalization results of orientation data in Chapter III (decoding high load on low load training data and vice versa), I pooled training data for the results presented here. However, I also performed training and testing separately (Figure IV-7) to be sure the result was truly comparable to those presented in the previous chapter. Overall, results remain largely comparable. However, likely due to the two-fold increased

number of training data, results from pooled training data looked more stable and were less volatile. Even when pooling training data, the study presented here still had fewer training examples than previously published results on attended motion which often include multiple scanning sessions per participant (Kamitani and Tong, 2005a, 2006; Serences and Boynton, 2007a). By pooling training data, I successfully increased training data without increasing scanning time. I therefore conclude that pooling training data, at least in this case, was a highly beneficial idea. More generally, I conclude that more training data leads to a more accurate voxel-selection and, in turn, to a more accurate decoding performance.

In this chapter, I have repeated the entire MVPD analysis twice, once with single volumes for training and testing and once with block averaged volumes. Even though not identical, both methods independently replicated very similar results. The block average method (Figure IV-6) seemed slightly superior in terms of actual accuracy but was also subject to higher volatility. The single volume method (Figure IV-7) was more robust, less volatile, but also slightly less accurate in the results. I speculate that with more data the block average technique would slowly become more consistent, additional to the benefit that it produced higher de facto classification accuracies. Further research has to yet determine this question in full detail. However, for this study, both methods were valid and comparable albeit single volume decoding seemed the more conservative approach.

*Load decoding.* I attempted to decode high vs. low load, however, MVPD was only marginally successful in decoding load from early visual cortex and V5/MT from single volumes (Figure IV-6, yellow bars). Additionally, when decoding from block averages, load decoding was less accurate and not significantly different from chance, even though decoding results for motion direction were more accurate (Figure IV-7, yellow vs. blue bars). Considering that decoding from single volumes seemed the more conservative approach, these two results leave an incoherent picture about whether load can or cannot be decoded by MVPD. My fourth hypothesis for the experiment presented in this Chapter was that load was decodable (similar to chapter III). With orientation data MVPD was sensitive enough to decode load significantly different from chance, in case of V1 and V2, even though the actual accuracies were less than 5% different from chance. For data from this chapter a similarly perplexing result emerged for MVPD with single volumes: V2, V3 and V5/MT show low accuracies but were significantly different from chance when examining the group result of all participants. Again the main effect of overall brain activity in an ROI seemed orthogonal for this prediction (main effect present in V1-V3, but not in V5/MT). Therefore, potentially, load could be spatially distributed in early visual cortex. Against this conclusion, however, stands the non-significant finding of load-prediction on block averages and a lack of a biologically plausible explanation (like a columnar organization for orientation of direction of motion). Thus, even combining results from Chapter III and IV, load decoding remains an unsolved phenomenon.

Even so, one thing this load-decoding result does represent was further evidence for my conclusion that voxels selected during MVPD were a poor representation of the overall BOLD signal. This appeared true irrelevant whether there was a BOLD main effect (V1-V3) or not (V5/MT).

*Direction of motion decoding.* Taking into account the successful decoding of motion direction with block averages (Figure IV-7) results in this chapter successfully replicated previous decoding results of motion direction (Kamitani and Tong, 2005a; Serences and Boynton, 2007a). The same is true for early visual cortices V1-V3 and the decoding with single volumes which was so far only done on other types of stimuli (Haynes and Rees, 2005b, 2005a; Haynes et al., 2007, Chapter III). However, surprisingly, decoding of motion direction was decreased in V5/MT for decoding with block averages and not significantly different from chance for decoding with single volumes. This is surprising given that V5/MT is known to be particularly motion sensitive and has been studied for over two decades (Zeki et al., 1991; Tootell et al., 1995; Tootell and Taylor, 1995; Heeger et al., 1999). Why does a motion selective area decode motion worse than early visual cortex? Other studies report a similar decrease in decoding performance for V5/MT (Kamitani and Tong, 2006; Serences and Boynton, 2007a, 2007b), thus this seemed to be a general feature of V5/MT decoding that, so far, lacks a definite explanation. Kamitani & Tong (2006) speculated that the decreased performance was due to the spatially smaller volume of V5/MT (e.g. compared to V1). This, so the argument, left less chance to find as many biased voxels than in a spatially larger

area of (i.e.) early visual cortex. Yet, if this explanation would be true, it is very hard to understand why classification of load (Figure IV-6, V5/MT yellow bar) was actually slightly better in V5/MT. Thus, here, this volume-speculation seemed less relevant than previously speculated. Yet, the volume speculation at least represents one of only very few actual theories about the meaning of MVPD results and their interpretation (Bartels et al., 2008). For the finding of reduced motion decoding in V5/MT, other possible theories might include: (i) motion selective cells in V5/MT might not be distributed in an anisotropic fashion, therefore MVPD was highly unlikely to indentify biased voxels. (ii) motion selective cells in V5/MT might be distributed but sub-ideal for the resolution I chose to scan with. Or (iii) motion was a distractor in our study and as such its perception in V5/MT was actively suppressed which decreased MVPD accuracy. While I cannot totally rule out any of the above speculations, it seems unlikely that any single one of them was particular dominant including the volume argument by Kamitani & Tong (2006). Yet, they might all be true to some extent, but there was no clear indication of a single reason being more relevant than another. Perhaps combinations of all factors lead to slightly noisier data in V5/MT that simply made it harder to decode motion direction.

Decoding of motion direction was successful under both load conditions in the central task. However, in none of the MVPD results was a significant difference between decoding accuracy of high and low load. In other words, comparing high and low attentional load conditions, classification accuracies did not differ across conditions in early visual cortices V1 to V3 and V5/MT. Thus, the neuronal

representation of motion direction of an irrelevant distractor, as measured by MVPD of the BOLD signal, in V1 to V3 and V5/MT remained unchanged by varying attentional load conditions of a central task. The neural representation of distractor motion direction, as measured by MVPD, was unchanged when participants performed a difficult (versus easy) task. Taken alone, this result was surprising given the strong and successful behavioral manipulation of the central load task (Figure IV-2) and the main effect of load in overall BOLD signal of V1 to V3 (Fig IV-5).

However, it is essential to discuss results from this chapter in conjunction with those from chapter III. In chapter III, the main result was that MVPD of distractor orientation was highly successful but remained unchanged under low or high load of the central task. Here, I investigated another basic feature, direction of motion, also known to be represented in a columnar architecture in the cortex. Results in this chapter found unchanged decoding accuracy for the visual feature direction of motion under high or low load in the central task, echoing results obtained for orientation. As a consequence, the speculation that findings of chapter III were limited to orientation seems less likely. On the contrary, taken together, I speculate that varying attentional load towards a central task, does not modulate how basic visual features are represented in early visual cortex V1-V3 and V5/MT. This was at least true for the basic visual features orientation and direction of motion as measured by MVPD of the BOLD signal.

Immediately the question of why the representation of these basic visual features was comparable under low and high load becomes apparent. However, before addressing the question about missing differences between load conditions it is important to consider how these successful decoding results were possible at all. To answer this more general question, I carefully considered what underlies the results observed, and ultimately, the BOLD signal in the voxel selection from ARD-SLR is relevant (automatic relevance detection for sparse logistic regression, for details, see Chapter II: General Methods). For orientation, I concluded that overall BOLD signal changes did not directly correspond to multivariate results, thus, the selected voxels did not reflect trends in the overall BOLD-signal. Load-decoding results from this chapter further corroborated this conclusion. Rather, I speculate that selected voxels represent brain activity that was (strongly) biased towards the basic visual feature that was tested for (here direction of motion). If this feature was represented in a biased way in some voxels, these were found by ARD and, hence, classification was successful. In the context of the results of this Chapter, some voxels in V1-V3 and MT unequally represent the motion directions I tested for, leading to a successful prediction. Why those unbalanced voxels exist can only be speculated. An anisotropic distribution of feature specific cortical columns is thought to underlie successful decoding of orientations (Haynes and Rees, 2005a; Kamitani and Tong, 2005b). And indeed, for V5/MT there is evidence of a columnar organization of direction of motion in monkey cortex (Albright *et al.*, 1984; Malonek *et al.*, 1994; Tolias *et al.*, 2001). Although direction-selective columns have not been demonstrated in early visual cortical areas of

monkeys, such columns have been reported in striate and extrastriate areas of other animals (Shmuel and Grinvald, 1996; Weliky et al., 1996). For monkeys and humans strong evidence for the existence of a direction-selective organization of visual motion processing comes from studies reporting reliable effects of direction-selective adaptation of the BOLD signal (Tootell et al., 1995; Nishida et al., 2003; Seiffert et al., 2003). Taken together, it seems relatively certain that the existence of biased voxels for motion direction was due to the existence and anisotropic distribution of direction selective columns in early visual areas and area V5/MT. This argument might also work vice versa: due to the fact that we find motion-direction selective voxels with ARD it might be likely that the underlying neuronal organization of direction of motion was columnar, or at least organized in a way allowing the existence of biased voxels.

Coming back to the question why there was no difference in the decoding performance of MVPD on such biased voxels, it seems that the bias (i.e. possibly more or less columns of one direction) present in these voxels was unchanged between the different load conditions. In other words, those voxels biased by the direction of motion in the distractor remained equally biased irrelevant of the central load condition. Additionally, result of this study showed that varying attentional load in a central task also had a BOLD main effect in V1-V3 (for orientation and direction of motion) but no BOLD main effect in MT for the irrelevant distractor. Thus, varying attentional load in the central task only modulated the overall BOLD response of the irrelevant distractor, but it did not change the

feature-specific (i.e direction of motion specific, or orientation specific) information of an irrelevant distractor contained in individual voxels.

This result must be considered carefully when trying to gain insights into the neuronal organization of human early visual cortex. By no means has it presented conclusive evidence for any specific neuronal organization; however it still represented a furthering of our understanding of the nature of attentional effects in the early visual cortex. For now I conclude that many behavioral measures as well as overall measures of the BOLD signal decreased under high load (Lavie and Fox, 2000; O'Connor et al., 2002; Pessoa et al., 2002; Jenkins et al., 2003; Lavie and de Fockert, 2003; Lavie, 2005; Schwartz et al., 2005; Lavie, 2006; Forster and Lavie, 2007, 2008). However, the underlying, detailed, visual feature specific representation of an irrelevant distractor might still remain relatively unchanged by varying attentional load towards a central task. This speculation would mean that the attentional demands of a central task might be somewhat irrelevant for this feature specific, neuronal representation. Considering the similar results for the basic visual features of orientation and direction of motion, these results might be replicable with other basic visual features such as color or maybe even luminance – of course so far limited to feature specific neural information as read out with MVPD. However, the underlying neural causes of this conclusion are relatively open and remain largely speculative.

## IV.5 Conclusion

Experimental results presented in this chapter lead to a number of conclusions. From the BOLD main effects of brain activity related with the irrelevant distractor (here coherently moving dots), I found a surprising lack of reduced activity in area V5/MT under high load. This was despite a variety of other studies reporting other measures reduced under high load, behaviorally and in terms of neural activity. However, the one classic study (Rees et al., 1997) that was most similar and relevant to my experimental design was actually reporting an interaction effect rather than a main effect. Therefore, I conclude that only a further experiment, exactly replicating this interaction, will be able to clarify whether or not this lack of a main effect in V5/MT was a non-replication of Rees et al. (1997) or simply an observation that was not mentioned in this classic study.

From the methodological point of view, I drew three conclusions. Firstly, for the comparison of MVPD with pooled training data and training data separately for the two load conditions, I concluded that it was: a) Beneficial for MVPD to have more training data as the result became more stable and less volatile. b) In cases analog to the one presented here, it might be a valid method to pool training data from multiple conditions to then test them separately. In general, more training data seemed to lead to a more accurate voxel-selection and a more accurate decoding performance.

Secondly, I concluded that the use of single volumes for training and testing in MVPD as well as the use of block averaged volumes were comparable techniques. Even though not identical, both

methods replicated very similar results. The block average method seemed slightly superior in terms of actual accuracy but was also subject to higher volatility. The single volume method was more robust and also less volatile in the results but produces slightly less accurate decoding.

Thirdly, the results for direction of motion decoding but also the results from load decoding showed that the voxel selection by ARD was a poor estimate of overall BOLD changes in an ROI. This was a confirmation of results from chapter III. Rather, I concluded that ARD selected voxels upon the classified feature only, irrelevant of the overall trend of brain activity in an area. MVPD was successful when there were voxels that show selectivity, for example caused by a columnar cortical architecture. I further speculated that this argument might be reversible: if biased voxels for a specific feature can be found, then this feature might imply cortically organization in a way readily allowing such biased voxels, as for example a feature-specific anisotropic columnar organization.

From the results of direction of motion decoding under high and low load, I described successful decoding of the direction of motion in V1 to V3 and V5/MT. This was in the context of attentional load being manipulated in a central task and visual feature specific MVPD measures considered from an irrelevant distractor. Further, I did not find a load effect between decoding results for high and low load, similar to what was described in chapter III for orientation. Taking the results for orientation and direction of motion together, I speculated that it might be possible for other basic visual features

of irrelevant distractors to be unchanged by attentional load conditions in a central task too. Lastly, I concluded that the underlying neural organization of this indifference of feature specific distractor activity towards attentional load in a central task remain largely unknown.

# Chapter V:

# Effects of attentional load on motion selective processing II

## V.1 Introduction

This chapter builds closely on findings and conclusions from the previous two, addressing questions that were opened up during analysis and interpretation of the results. In chapters III and IV, I presented results from attentional load paradigms that are not immediately interpretable in a straightforward way by load theory (Lavie et al., 2004). In short (for detailed description see chapter I: General Introduction), load theory describes attentional resources as limited and concludes that a demanding, high load, task consumes more attentional resources than an easy, low load task. Consequently, many studies find measures related to non-task-relevant stimuli, so called irrelevant distractors, decreased under high load as fewer attentional resources are available to process them. The classic load-study describes such effects on distractors consisting of color-shape conjunctions (Lavie, 1997), but the variety of measures, including irrelevant distractors, affected by attentional load is large (for details see chapter I, examples include: Rees et al., 1997; Pessoa et al., 2002; Jenkins et al., 2003; Lavie and de Fockert, 2003; Jenkins et al., 2005; Schwartz et al., 2005; Forster and Lavie, 2007, 2008; Torralbo and Beck, 2008). Conversely, in chapter III, multivariate decoding accuracy of the basic visual

feature orientation in early visual cortex remained unchanged by varied attentional load demands in a central task. Chapter IV presented a qualitative replication of such a multivariate decoding result, this time for the basic visual feature of direction of motion. The counterintuitive result of equal decoding ability under low and high attentional load (towards a central task) was thus reproduced with distractors containing two independent, basic visual features. As a consequence, one of the main conclusions of chapter IV was attentional load towards a central task is likely to be irrelevant to the decoding of basic visual features from early visual areas in general. While showing multivariate decoding results different from chance under both high load and low load, at the same time, chapter IV left two critical points unaddressed: a) differences in eye movements between conditions might have influenced results; b) early visual areas V1 to V3 showed an (expected) main effect of decreased distractor-related brain activity under high load, but V5/MT did not show a significant reduction. In other words, the previous study did not show a difference in mean activity between high load motion and low load motion. This is surprising because, Rees et al (1997) reports an interaction between brain activity related to motion vs. no motion under low vs. high load (Rees et al., 1997). However, the two studies were not directly comparable due to the fact that the experimental design in chapter IV lacked two "no motion" conditions (under high and low load) and, thus, an interaction term could not be examined.

### V.1.1 Objectives of this study

The present study presents an experimental paradigm building on the experimental paradigm of chapter IV, but now addressing the concerns raised above. Firstly, eye movements were recorded continuously for 4 participants during the fMRI scan. Secondly, the experimental paradigm was extended by two additional no motion conditions containing only static dots as a distractor in order to be better comparable with previous studies (Rees et al., 1997).

### V.1.2 Hypothesis

My hypotheses for the experiment presented in this chapter were: Firstly, I hypothesized that I would replicate the main findings from chapter IV: a main effect of decreased BOLD signals under high (vs. low) load in V1 to V3, and, for V5/MT, no such significant local reduction of brain activity. Further, I expected to find multivariate decoding results parallel to those in chapter IV where no difference between distractor motion decoding under high load and low load could be found.

Secondly, I hypothesized that the addition of the "no motion" conditions would enable me to replicate an interaction of brain activity related to motion vs. no motion under low vs. high load (Rees et al., 1997).

Finally, my last hypothesis was that eye-movements would be near identical between the different experimental conditions. I will thus test for any differences between eye-movements during the different conditions. If no differences should be found, this could be seen as an indication towards excluding eye-movements as potential confounding factor.

## V.2 Methods

### V.2.1 Participants

Eight healthy participants (4 male, mean age 28 years) gave written informed consent to participate in the study, which was approved by the local ethics committee. All participants had normal or corrected to normal vision and were new to the task. One subject was excluded from analysis due to excessive head movements which made it impossible to coregister all experimental runs.

### V.2.2 Stimulus

During the main fMRI experiment, participants performed a visual detection task on centrally presented stimuli identical to the one described in chapter III and IV. In brief, participants had to respond either to a single colour or a conjunction of two colours and orientations depending on instructions. This made the task either demanding (conjunction task - high attentional load) or relatively

effortless (colour task - low load). The irrelevant distractor motion was identical to that described in chapter IV, 100% coherently moving dots contained in an annulus shaped region around the central task. Additionally, two experimental stimuli were introduced, containing only static dots in the same annulus than before. These dots remained unchanged during the whole duration of the experimental block. Combined with two different attentional load instructions, these static dots resulted in the two new conditions: static low load and static high load.

### V.2.3 Procedure

Equipment and software used were identical to those described in chapter III and IV. Additionally, eye position was continually sampled for 4 participants at 60 Hz using an ASL 504 LRO infrared video-based MRI compatible eye tracker (Applied Science Laboratory, Bedford, MA).

In all participants the block length of every motion and no motion block was extended to 30 seconds each to match experimental conditions of a previous study (Rees et al., 1997).

For the 4 participants that had eye movements measured during the experimental block, the arrangements of blocks within sessions was altered: These 4 participants were scanned for a total of 48 short experimental sessions. Each session consisted of exactly one experimental block preceded and followed by a short rest period.

This resulted in eight sessions per conditions, each with one experimental block. Motion direction/ static dots and difficulty of the central task were counterbalanced across sessions and participants. Thus, in detail, each session started with the appearance of the central fixation spot for 8s including a short reminder of the task to come, followed by a 30s long block of the load task at central fixation with the motion stimulus/ static dots surrounding the central task (Figure IV-1 A). The last 20s of each run were used to give participants a short rest and only comprised a fixation cross. Each experimental run thus lasted 48 volumes.

The other 3 participants underwent eight sessions containing each of the six conditions once per session, equally summing up to 48 experimental blocks. Motion direction/ static dots and difficulty of the central task were counterbalanced within sessions.

### V.2.4 Eye tracking Analysis

To evaluate the possibility that a difference in eye movements between different conditions might be able to explain any differences found in brain activity, I compared the mean and standard deviation of eye movements along the X and Y axis and the mean and standard deviation of pupil size. This was done as to test for differences between the conditions. To this end, I first divided the X, Y eye position and pupil size values of each run by its mean (normalization) and then calculated the standard deviation of these values. Thus, I computed a distribution of 32 standard

deviations per experimental conditions (8 runs per condition, per subject). I evaluated each of these 6 condition specific distributions for any differences with a separate one-way analysis of variance (ANOVA). Finally, I specifically tested for a difference between motion runs under high load and low load with a paired two-sample t-test.

### V.2.5 Scanning

Imaging parameters for the main experiment and stimuli for meridian mapping were identical to those described in chapter III. To identify motion sensitive area V5/MT, an V5/MT-localizer sequence was applied. Stimuli and imaging parameters for the V5/MT localizer were identical to those described in chapter IV.

### V.2.6 Analysis

*Data preprocessing.* All data were preprocessed using Statistical Parametric Mapping software (SPM5, http://www.fil.ion.ucl.ac.uk/spm). The first 5 volumes of each fMRI scan were discarded to allow for magnetic saturation effects. The remaining functional images volumes were realigned to the first image, and then the structural scan of each participant was co-registered to their functional data. Functional data of the main experiment were not spatially smoothed, but data from the meridian mapping and V5/MT localizer runs were spatially smoothed with a Gaussian kernel of 5mm full-width half-maximum. Thus, data

preprocessing, was identical to methods described in chapter III and IV.

*ROI localization*

Different to chapter III and IV, I utilized FreeSurfer (http://surfer.nmr.mgh.harvard.edu) for segmentation and cortical flattening using each subject's specific structural image that was previously coregistered to all functional data. FreeSurfer is thought to result in much higher quality reconstructions of structural data, thus we found it prudent to change from previous techniques to this new, presumably better, 3D-reconstruction software package (for details see chapter II). Standard meridian mapping procedure was employed to identify the borders of the visual areas V1, V2 and V3 in the occipital cortex (Sereno et al., 1995). V5/MT was identified overlaying the contrast motion vs. no motion onto the surface. In order to extract activity from these ROIs, I created mask volumes for each ROI. Given the high degree of accuracy of the standard FreeSurfer segmentation, these mask volumes should contain gray matter voxels only.

*Timecourse analysis and univariate analysis*

To compute the percent signal changes in the time course analysis, I used the mean raw activation of the realigned timecourse correcting only for slow signal drift typical in fMRI scanning by high-pass filtering (cut-off – 128s). Each session was then scaled by its own mean activity during the entire session. This linear transformation was repeated for each ROI and each subject separately. Thus, the resulting time series were sorted and

averaged in each ROI according to four conditions: motion under high load; motion under low load; no motion under high load; and no motion under low load (Since upwards left and upwards right motion did not result in a differential univariate result, they were collapsed).

I then used SPM5 to perform a within-participant analysis, using a voxel-wise general linear model (GLM) that comprised six delayed boxcar waveforms representing the six experimental conditions. During this analysis the fMRI time series were high-pass filtered (cut-off – 128s) and global changes in activity were removed by proportional scaling of each session. Apart from the new static conditions, the only difference in comparison to the analysis in chapter IV was a slight adjustment in the timing variables of the GLM according to the extended 30 second experimental blocks. The resulting regressor images (one for each box-car regressor), represented brain activity effect size at each particular voxel in percentage units of the whole brain mean, and convolved by the standard hemodynamic response (HDR) function. For each condition these regressor-images were averaged, then extracted across all voxels in each ROI, yielding a single value of percent-signal change. This value represented the percent-signal change of this ROI, relative the global brain mean. Thus, this procedure enabled me to directly compare signal change for each experimental condition in V1, V2, V3, and V5/MT for each subject. As in chapter IV, I collapsed across different motion directions to obtain an equal amount of measurements for all conditions.

Since all condition values were relative to the same global brain mean activity per subject, a repeated measurement ANOVA, where each subject is a new repetition, was used to compare for differences of brain activity between conditions. Additionally, to compare results directly to chapter IV, I compared pairs of motion conditions (motion under high and low load per subject) with a paired t-test.

Finally, to compare results presented here to reports from a classical study, I directly tested for an interaction of motion (vs. no motion) greater under high (vs. low) load (as reported by Rees et al., 1997). This was done by performing a paired t-test between the differences of motion (high load) minus static (high load) against the differences between motion (low load) minus static (low load). Pairs of differences were always defined as values from the same subject.

*Multivariate pattern classification.*

Multivariate pattern analysis was identical to chapter IV, and is thus only briefly described (for detailed description see chapters II, III and IV): Unsmoothed, realigned fMRI data from the experimental runs were adjusted for the lag in hemodynamic response function by shifting all block-onset timings by 5 volumes (6½ seconds). To determine classification accuracy, only classification with unseen and independent-test data was considered. Pattern classification was performed using a sparse logistic regression (SLR) algorithm (Yamashita et al., 2008). I pooled training data from both load conditions. Thus, I trained the SLR algorithm with double the

number of upwards-left versus upwards-right examples, irrespective of the load condition of the central task. During testing, however, I kept-testing data strictly separate for low and high load, resulting in two different decoding results achieved upon the same training data. Significant differences from chance (2 categories = 50% chance) were tested for with Student's one-sample t-tests, applying Bonferroni correction for multiple comparisons across high and low load and all ROIs examined.

In order to pool multivariate results from chapter IV and results form this chapter, only the first 10 volumes per block were considered for multivariate analysis. However, when analyzing the entire block length instead, results were virtually identical, confirming that the pattern of decoding accuracies was consistent throughout the block.

In order to evaluate the probability that the classification was driven by over-fitting of arbitrary patterns of spatial correlations in the data, I carried out a shuffle-control test (Mur et al., 2009). If the assumption that classification is driven by chance were true, similar results should be obtained if labels indicating the experimental condition for each example vector were shuffled randomly. To test this, I ran a separate analysis where labels of the test examples were re-shuffled for each round of the cross-validation procedure. The resulting distribution of classification accuracy characterized the expected distribution of accuracy under the null hypothesis.

## V.3 Results

### V.3.1 Behavioral results

In both load conditions, performance was always above chance, confirming adherence to the task requirements. I examined Reaction Times (RT) across participants comparing the RTs of low load conditions with the RTs of high-load conditions. Mean detection latencies for central target crosses were significantly slower in the high-load versus low-load condition [Reaction Times High Load: 608ms; Low Load: 485ms; $t(47) = 4.1$, $P < 0.001$] (Figure V-1, left). Hit rates significantly decreased in the high load condition [Correct Responses Low Load: 97.6%; High Load: 86.6%; $t(47) = -12.78$, $P < 0.001$] (Figure V-1, right). There also was a significantly increased number of false alarms during the high load task [Error Rate High Load: 31.3%; Low Load: 13.3%; $t(47) = 5.3$, $P < 0.001$]. Together, these results indicate that the central task successfully manipulated load.

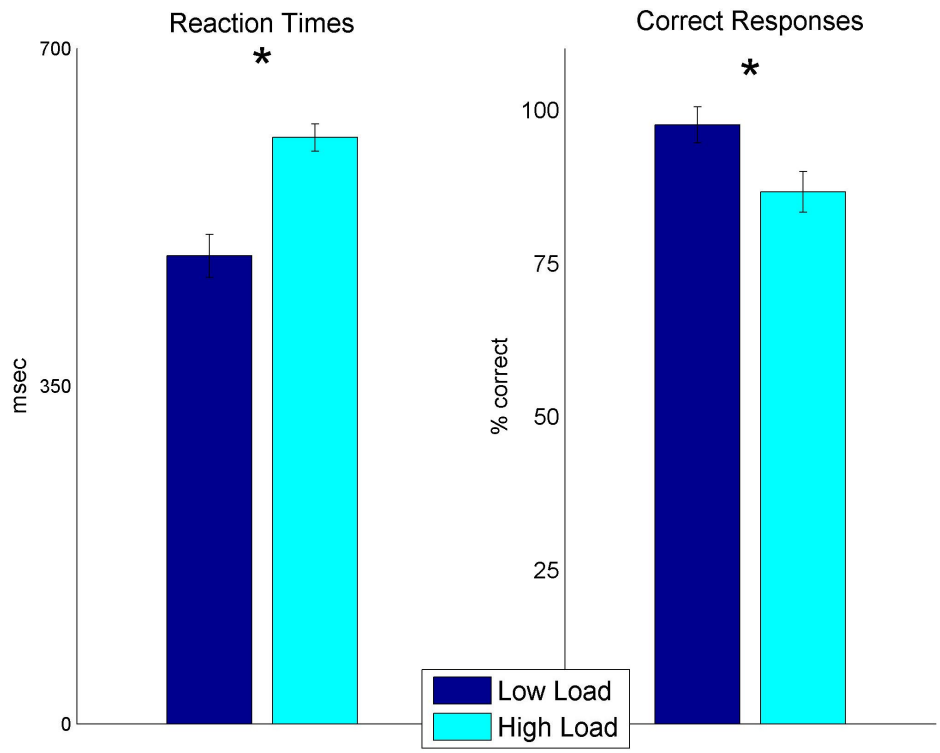



**Figure V-1**
**Shown are behavioral results of performance during the load task. Mean reaction times of subjects for successful detection of target crosses are shown on the left. Mean correct hits across subjects for target crosses are displayed on the right. Separate colors denote low (dark blue) and high (light blue) load conditions in the central task. Both differences are highly significant (P< 0.001) and present in every subject. Errorbars indicate standard error of the mean (SEM) across scans.**

### V.3.2 Eye tracking results

For four participants, eye tracking data that were sampled continuously throughout experimental blocks. This part of the analysis is new in comparison to chapter IV. In depth inspection of per-block mean and standard deviation of eye position and pupil size showed that there were no significant differences between conditions. A separate one-way analysis of variance (ANOVA) for X,

Y positions and pupil size did not show any significant differences between eye movements obtained from the six experimental conditions [movement along X axis: Means = (0.1014; 0.1066; 0.1030; 0.0876; 0.1139; 0.0901), F(186) = 0.23, p = 0.95; movement along Y axis: Means = (0.1619; 0.1696; 0.1767; 0.1683; 0.1837; 0.1590), F(186) = 0.19, p = 0.96; Pupil size, Means = (0.1778; 0.1816; 0.1558; 0.1646; 0.1630; 0.1663), F(186) = 0.26, p = 0.94]. Finally, a specific two-sample t-test-testing for a difference between motion runs under high load and low load, again, showed no significant differences [movement along X axis: t(126) = 0.42, p = 0.66; movement along Y axis: t(126) = -0.34, p = 0.73; Pupil size: t(126) = 1.03, p = 0.3].

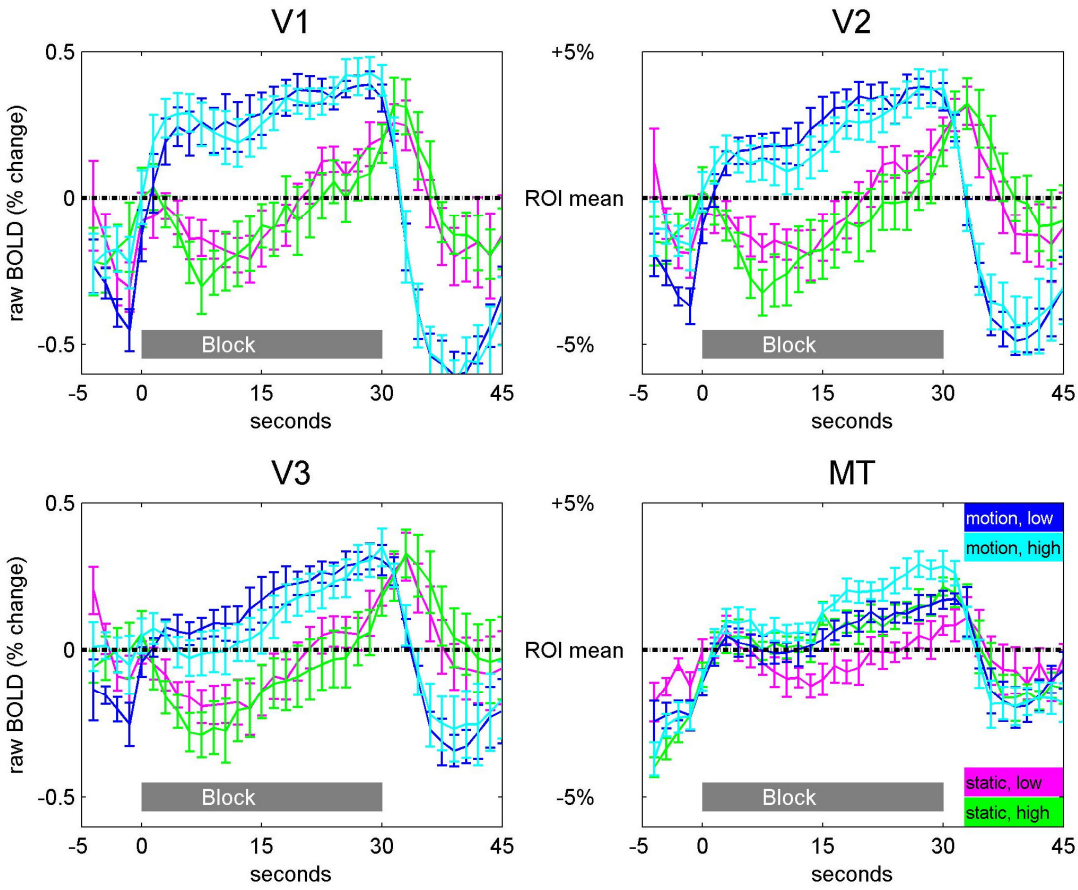



**Figure V-2**
**Timecourses of the mean BOLD signal changes from all ROIs. Activity of blocks of all experimental conditions was averaged over all 7 participants. The Y axis illustrates the percent signal change in comparison to activity to the session mean within each ROI. The X axis depicts time in seconds aligned to each block onset. Each stimulus block was 30 seconds long, indicated by the gray bar. All ROIs showed a clear increase of BOLD activity shifted by the hemodynamic delay. While, in V1 to V3, activity during the static conditions decreased below the ROI mean, in V5/MT, such a decrease in activity was only visible for static under low load.**

### V.3.3 Univariate results

I first plotted the timecourse of raw BOLD signal change for each subject and in each ROI. Figure V-2 displays the mean timecourse of raw BOLD signal change. Time zero marks the onset of the 30 second long blocks. In all ROIs the BOLD signal associated with all

conditions reached an initial peak about 6-8 seconds after block onset, consistent with the timing of the hemodynamic response. For motion conditions, after the initial activation peak, motion related activity remains increased and even increases further during the duration of the block. Contrary, in V1 to V3, activity during the no motion/ static conditions decreased below the ROI mean after the initial peak and only reaches a second peak at the end of the block. However, in V5/MT, such a decrease in activity was visible for static peripheral dots under low load for a lesser degree; under static high load activity remained above ROI mean activity during the experimental block.

To quantify these raw signal changes, I explored mean activity within each ROI obtained from the estimated regressor images from the GLM. Importantly, these values represent brain activity effect size in percentage units of the whole brain mean (different to the raw timecourse effects in Figure V-2 which represent signal change in percentage units of the ROI mean). Figure V-3A displays these mean signal changes for each condition in each ROI. To quantify differences within these results, I performed a repeated measures ANOVA. It reveals a significant main effect of motion in all ROIs, as well as a significant main effect for load in V5/MT [V1: Motion: $F(6) = 33.9$, $p < 0.01$, Load: $F(6) = 0.93$, $p = 0.37$, Motion x Load: $F(6) = 0.36$, $p = 0.57$; V2: Motion: $F(6) = 79.8$, $p < 0.01$, Load: $F(6) = 4.13$, $p = 0.09$, Motion x Load: $F(6) = 1.32$, $p = 0.29$; V3: Motion: $F(6) = 110.5$, $p < 0.01$, Load: $F(6) = 5.84$, $p = 0.052$, Motion x Load: $F(6) = 3.01$, $p = 0.13$; V5/MT: Motion: $F(6) = 10.7$, $p =$

0.02, Load: F(6) = 38.31, p < 0.01, Motion x Load: F(6) = 1.59, p = 0.25].

However, more importantly, here the experiment was designed in order to extend and clarify surprising findings from chapter IV. In chapter IV, I report a main effect of decreased local brain activity in V1 to V3, and, for V5/MT, no such significant local reduction of brain activity. Figure V-3B shows this effect, equivalent to Figure IV-4 (identical color-code): I subtracted activity of motion under high load from activity from motion under low load. V1 to V3 all show similar tendencies towards greater activity under low load, however, this difference is only significant in V2 and V3. In V1 the result is only significant if Bonferroni correction (n = 4) for multiple comparisons is not applied. Apart from the result in V1 this is a qualitative replication of results in V1 to V3 in chapter IV [V1: Low load motion: 0.54, High load motion: 0.43, t(13) = 1.8, P = 0.18; V2: Low load motion: 0.54, High load motion: 0.33, t(13) = 3.6, P < 0.01; V3: Low load motion: 0.33, High load motion: 0.093, t(13) = 4.1, P < 0.01].

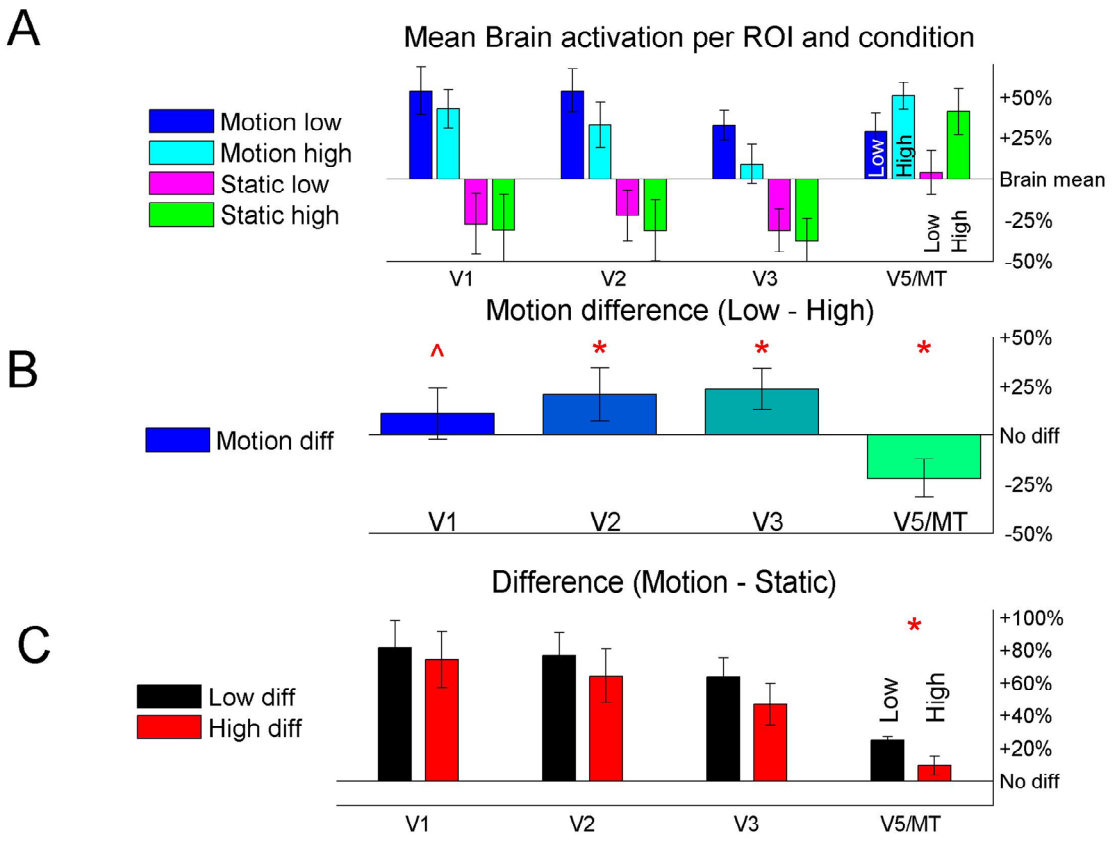



**Figure V-3**
**A) Shown are mean brain activation values per ROI and experimental conditions compared to the whole brain mean activation. For each ROI, the graph plots the mean brain activation during the experimental blocks of distractor motion under low attentional load (dark blue), distractor motion under high load (light blue), distractor static dots under low load (magenta) and distractor static dots under high load (green). Errorbars denote inter-subject SEM (n = 7).**
**B) The graph is showing the difference in mean activity in each area comparing high and low load, averaged across scans and subjects. Values are the mean brain activity of high load blocks subtracted from the mean low load activity. While activity is significantly reduced under high load for V2 to V3 (* = p < 0.01), this difference is only significant in V1 when not correcting for multiple comparisons (^ = p < 0.05 uncorrected). In V5/MT there is a significant activity difference in the other direction: activity under high load as compared to low load is increased. Errorbars denote inter-subject SEM (n = 7).**
**C) The graph shows the difference of motion minus static activation under low load (black) and the same difference under high load (red). What can be easily observed is a large difference between activity related to motion and static under low and high load in the central task in all ROIs. The general trend towards reduced activity under high load conforms with Load theory (Lavie et al., 2004). In V5/MT this difference becomes significant (* = p < 0.01), representing a direct replication of pervious work (Rees et al., 1997). Thus, this resolves the perceived contradiction from chapter IV. Errorbars denote inter-subject SEM (n = 7).**

For V5/MT, activity in the motion conditions was significantly decreased under low (vs. high) load. This is a qualitative replication of results in chapter IV [MT: Low load motion: 0.29, High load motion: 0.51, $t(13) = -4.5$, $P < 0.01$].

Finally, the additional no motion/ static conditions allowed me to repeat a similar interaction analysis to that previously reported (Rees et al., 1997). Rees et al. (1997) reported an interaction between load (low vs. high) and motion (motion vs. static) such that brain activity for motion (vs. no motion) was increased under low (vs. high) load in V5/MT. Figure V-3C depicts the difference between motion and static conditions under high and low load. A paired t-test reveals that these differences were significantly different only in V5/MT. Thus, activity of motion (vs. no motion) is significantly higher under low (vs. high) load. Consequently, I directly replicated previous results in V5/MT [V1: Motion minus Static (Low load): 0.77, Motion minus Static (High load): 0.79, $t(6) = -0.09$, $P = 0.93$; V2: Motion minus Static (Low load): 0.68, Motion minus Static (High load): 0.73, $t(6) = -0.3$, $P = 0.77$; V3: Motion minus Static (Low load): 0.55, Motion minus Static (High load): 0.57, $t(6) = -0.17$, $P = 0.87$; MT: Motion minus Static (Low load): 0.33, Difference, Motion minus Static (High load): 0.013, $t(6) = 4$, $P < 0.01$].

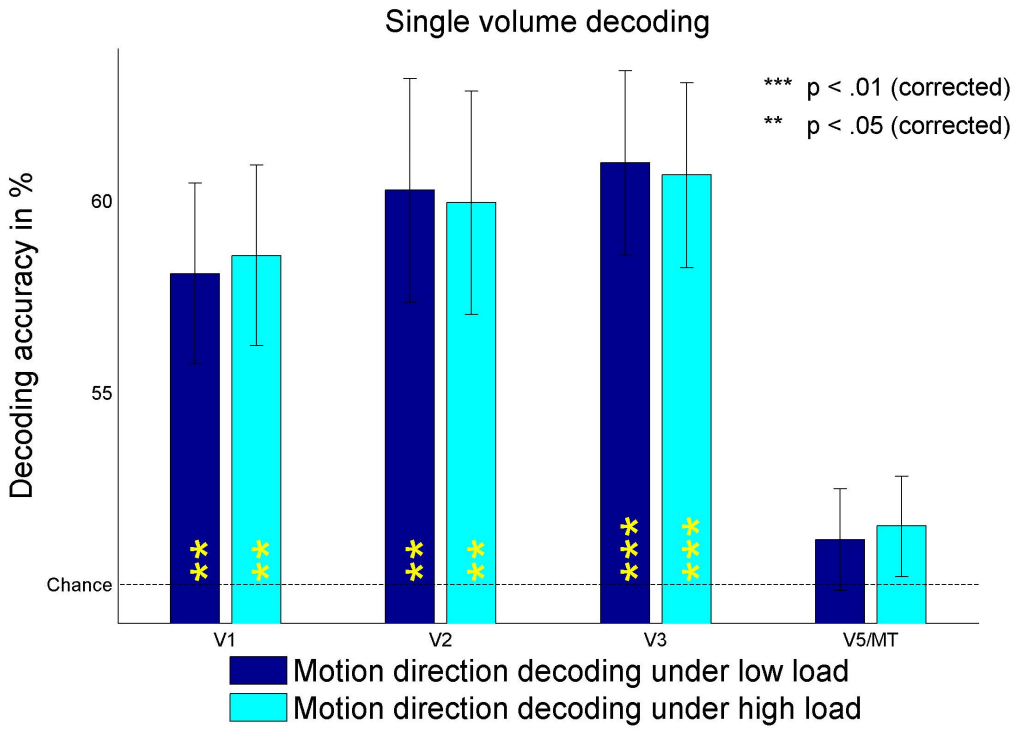



**Figure V-4**
**Decoding accuracy (mean across 13 subjects) for visual areas V1 to V3 and V5/MT, obtained with Sparse Logistic Regression (SLR). Different colors indicate decoding performance for different test-vectors: left-upwards vs. right-upwards (distractor-) motion decoding was successful in most ROIs, compared to chance, under both low load (dark blue bars) and high load (light blue bars) conditions. Yellow stars indicate results significantly different from chance after Bonferroni correction: *** = p < .01, ** = p < .05 (one tailed t-tests). In both conditions, training data was pooled across load conditions to increase power of training. A further paired t-test between the two orientation decoding results (per ROI) revealed no significant differences in any area. Errorbars denote SEM across subjects.**

### V.3.4 Multivariate classification result

Like in previous chapters (III and IV), I report classification results obtained with SLR- MVPD. Figure IV-4 shows decoding accuracies for areas V1 to V3 and V5/MT from all participants in chapter IV and V (n = 13). The y-axis depicts mean accuracy of classification across participants, while different colors along the x-axis indicate

decoding from different experimental conditions in the different visual areas. Training data was pooled to increase training power (see Methods). Decoding left-upwards versus right-upwards direction of motion was significantly different from chance in V1 to V3 under high and low load. However, in V5/MT, prediction accuracy was closer to and not significantly different from chance. All results have been Bonferroni corrected for multiple comparisons (2x4 ROIs) [V1: Low Load: 57.8%, $t(12) = 3.3$, $P = 0.02$; High Load: 58.1%, $t(12) = 3.3$, $P = 0.02$; V2: Low Load: 59.7%, $t(12) = 3.3$, $P = 0.02$, High Load: 59.4%, $t(12) = 3.2$, $P = 0.03$; V3: Low Load: 60.6%, $t(12) = 4.4$, $P < 0.01$, High Load: 60.3%, $t(12) = 3.9$, $P < 0.01$; V5/MT: Low Load: 51.2%, $t(12) = 0.88$, $P = 1.58$, High Load: 51.4%, $t(12) = 1.1$, $P = 1.2$]. The difference between the two load conditions was not significant in any ROI [V1: $t(24) = -0.082$, $P = 0.93$; V2: $t(24) = 0.082$, $P = 0.94$; V3: $t(24) = 0.092$, $P = 0.93$; V5/MT: $t(24) = -0.14$, $P = 0.89$].

For completeness, I repeated the multivariate analysis in this chapter using only vectors that contained an average of always 3 volumes (TR). This approach closely mirrored other 3TR-averaging strategies as applied by some recent studies using MVPD (Kamitani and Tong, 2005a, 2006; Serences and Boynton, 2007a, 2007b). These studies successfully distinguished between different attended motion directions, especially also in V5/MT. However here, studying unattended motion direction, I found no significant improvement in decoding accuracies using this approach [uncorrected results of t-tests: V1: Low Load: 61.8%, $t(12) = 4.1$, $P < 0.01$, High Load: 55.1%, $t(12) = 1.7$, $P = 0.06$; V2: Low Load: 60.4%, $t(12) = 2.4$, P

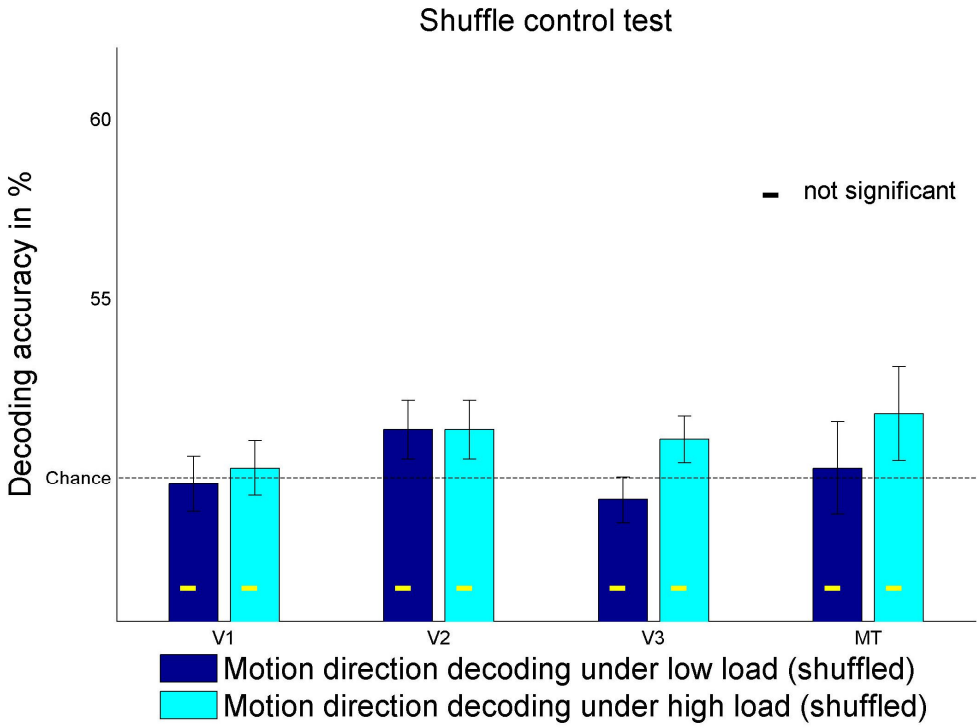



**Figure V-5**
**Results from randomly permuting labels for each set of test vectors in each cross-validation. No ROI shows any above chance prediction. The results of the permutation test confirm the distribution of classification accuracy expected under null hypothesis for each ROI.**

= 0.02, High Load: 62.1%, t(12) = 2.4, P = 0.02; V3: Low Load: 63.3%, t(12) = 2.9, P < 0.01, High Load: 62.1%, t(12) = 3.2, P < 0.01; V5/MT: Low Load: 55.3%, t(12) = 1.4, P = 0.10, High Load: 55.1%, t(12) = 1.2, P = 0.12]. Thus, overall the results from classification with block averages were highly comparable to single volume decoding. This closely mirrored similar findings in chapter IV.

To assess the possibility that classification was driven by stimulus unrelated spatial correlations in the data (independent from the direction of motion presented), I carried out a shuffle-control test

(Mur et al., 2009). To do this, I repeated the classification from all ROIs, but this time with shuffled labels for the test examples. Figure V-5 depicts the results of this shuffle-control analysis. Training the classifier using the same training sets but with shuffled labels for example vectors confirmed the distribution of classification accuracy expected under null hypothesis. This control analysis strengthened the main findings, as it reconfirmed the validity of the result and the independence of the data used to obtain them [Bonferroni corrected results, V1: Low Load: 49.8%, t(6) = -0.17, P = 4.52, High Load: 50.3%, t(6) = 0.18, P = 3.44; V2: Low Load: 51.3%, t(6) = 1.2, P = 1.09, High Load: 51.3%, t(6) = 1.3, P = 1.03; V3: Low Load: 49.4%, t(6) = -0.71, P = 5.98, High Load: 51.1%, t(6) = 0.75, P = 1.93; V5/MT:Low Load: 50.3%, t(6) = 0.15, P = 3.54, High Load: 51.8%, t(6) = 1.6, P = 0.669].

## V.4 Discussion

The experimental paradigm presented in this chapter is in essence an extension of the experiment presented in chapter IV. Thus, where applicable, all findings are directly compared with the corresponding results from chapter IV.

*Behavioral results.* Performance in the central task matched performance during in chapter IV and III. Participants successfully paid attention to the central task and central attentional load was effectively varied by our task manipulation, similar to previous applications of the task (Schwartz et al., 2005).

Eye-movements were not significantly different between conditions. Thus, eye tracking results from this chapter made the possibility that different results between conditions could be explained by differential eye movement behavior between conditions less likely. Because participants needed to keep their eyes at fixation in order to perform the RSVP task, this result was anticipated. However, it is reassuring to observe that, when measured, there is virtually no difference between eye movements in the different experimental conditions.

*Univariate results.* Our results provide insight into the seemingly contradictory findings in chapter IV and an earlier study by Rees et al. (1997). Specifically, chapter IV demonstrated a main effect of increased V5/MT activity under high load (vs. low load), while, at the same time, Rees et al. (1997) reports an interaction effect of decreased motion (vs. no motion) activity under high (vs. low) load. Reassuringly, the addition of the two static / no motion conditions allowed me to replicate both these findings. Activity in V5/MT did indeed increase under high load (chapter IV, Figure V-3B), however, when extracting the *motion-specific* component within this (increased) activity by subtracting motion unrelated activity obtained in the static conditions under high and low load, I found a decrease of motion-specific activity (Figure V-3C). Thus, I replicated the result of Rees et al (1997). In summary, while activity under motion high (vs. low) load increased, it increased *more* under static high (vs. low) load (activity bars for V5/MT, Figure V-3A). I conclude that the seemingly contradictory findings in chapter IV and

Rees et al. (1997) were in fact due to methodological differences. When the exact methodological approach of both studies was applied (current chapter), both results, chapter IV and Rees et al. (1997) are replicated.

In early visual cortices V1 to V3, motion-distractor related activity was significantly decreased under high load comparing activity in the motion conditions only. This trend could also be observed in the ROI-specific timecourses depicted in Figure V-2. To my knowledge this is the first demonstration that, irrelevant, distractor-motion related activity in V1 to V3 has been found reduced due to increased load in a central task. Other studies observing other types of distractors under differing load conditions, report similar results of reduced distractor activity in early visual cortices due to increased load in an attended task (O'Connor et al., 2002; Jenkins et al., 2003; Schwartz et al., 2005).

*Multivariate decoding results*. Multivariate pattern analysis allowed accurate classification of the neural representation of one of two motion directions from early visual cortices V1 to V3. Previous studies have reported similar findings for attended motion directions (Kamitani and Tong, 2005a, 2006; Serences and Boynton, 2007a, 2007b). Here we extend these findings to unattended, irrelevant motion. This extension is valid under high and low load in the central task. Comparing high and low attentional load conditions, decoding accuracies are roughly similar across conditions in early visual cortices V1 to V3. Thus, in V1 to V3, decoding of BOLD signal of distractor motion direction was not significantly different when

participants performed a difficult (versus easy) task at fixation. This result was found reliably in V1, V2 and V3 and stands in contrast to the univariate result in these ROIs (significantly decreased activity under high load). It mirrors results obtained for orientation specific activity in chapter III and substantially strengthens the conclusion that overall BOLD signal changes do not directly correspond to the representation of feature specific activity patterns as measured by MVPD.

For area V5/MT it was not possible to decode which direction of distractor motion was present during an experimental block. It is possible that the current experimental design failed to produce successful motion direction decoding from V5/MT due to the differences between previous studies and the current experiment: While previous studies studied *attended* motion directions specifically (Kamitani and Tong, 2005a, 2006; Serences and Boynton, 2007a, 2007b), our result is based on *ignored*, irrelevant motion. Additionally, it seems that even when attending motion direction specifically, decoding seems to be generally less successful from V5/MT (compared to V1 to V3, see for example Serences and Boynton, 2007a, 2007b). Chapter IV also discusses other potential reasons for this decoding-performance decline (small size of V5/MT; possible non-anisotropic distribution of motion selective cells in V5/MT; sub-optimal scanning resolution). For the current chapter, methodological details in 4 participants could have further impaired decoding performance due to imprecise coregistration of all runs (see below). Most likely all the above reasons combined led to the

non-significantly different from chance motion-direction decoding in area V5/MT.

*Experimental paradigm and analysis*. For 5 participants, I changed my experimental design to 48 single sessions with one experimental block each instead of 8 sessions of 6 blocks. While this was perfectly suited to facilitate eye tracking as the eye tracker could be recalibrated after each session, analysis of fMRI data from these participants proved more difficult than normal. This was due to the fact that preprocessing, especially coregistration, of experimental and localizer scans proved more difficult than expected. One subject had to be excluded simply due to the fact that, due to head-movements in-between scans, experimental data seemed impossible to realign for all 48 experimental scans. Additionally, MVPD critically depends on the assumption that any voxel contains the same underlying neuronal population in all volumes. No smoothing of data is performed. Thus, perfect realignment is absolutely vital for any result to be meaningful. The current study consisted of 48 independent, short scans. Within each scan there was one block of one of the 6 experimental conditions. That meant that, on average, the same experimental condition would only be repeated every 6 scans. In hindsight, this experimental design gave participants the additional opportunity to move 47 times, between every session, not even considering head-movements within the scan. Finally, the constant short sessions were reported as very tiring and demanding, making participants potentially move even more between scans (i.e. to stretch or in order to stay alert). In conclusion and retrospectively, given these obstacles it is surprising

that it was possible at all to analyze 4 out of 5 participants without any further problems.

## V.5 Conclusion

The experimental results presented in this chapter lead to a number of conclusions. Firstly, I demonstrated that differential eye movements between different experimental conditions are highly unlikely to be able to explain any of the results.

For brain activity associated with irrelevant moving dots, I find a reduction of overall signal strength in V1 to V3. This main effect of reduced activity under high load in early visual cortices is shown here for the first time for a motion stimulus. While it is the first demonstration of reduced motion related BOLD activity under high load, it mirrors similar results obtained using other distractor stimuli (O'Connor et al., 2002; Jenkins et al., 2003; Lavie, 2005; Schwartz et al., 2005). For V5/MT, I qualitatively replicated an interaction effect of a classic load study (Rees et al., 1997). Thus, I resolved the perceived discrepancy between this study and results presented in chapter IV.

From the results of direction of motion decoding under high and low load, I described successful decoding of the direction of motion in V1 to V3. This is in the context of attentional load being manipulated in a central task and visual feature specific MVPD measures considered from an irrelevant distractor. As in chapter III

for the basic visual feature of orientation, a load effect within decoding results for direction of motion remains elusive. Taken the results for orientation and direction of motion together, I speculate that other basic visual features of irrelevant distractors to be unchanged by attentional load conditions in a central task too. However, I conclude that the underlying neural organization of this indifference of feature specific distractor activity towards attentional load in a central task remains largely elusive.

For V5/MT, direction of motion decoding is unsuccessful. I speculate that it might be due to a number of reasons of which it seems most likely that direction of motion decoding failed due to the fact that motion was unattended and irrelevant in the current study.

# Chapter VI:

# The independence of feature-based attentional modulation and the representation of a behavioral decision in early visual cortices

## VI.1 Introduction

Conscious perception of visual information depends on neural activity at many levels of the visual system. However, to effectively process visual information humans have to select a small subset of stimuli to attend at each point in time. When multiple stimuli are simultaneously present in a scene, they thus compete for cortical representation and access to awareness (Desimone and Duncan, 1995; Serences and Yantis, 2006). It is widely believed that humans resolve this competition by selectively filtering incoming sensory input, based on current behavioral goals so that relevant stimuli are processed more efficiently than irrelevant stimuli. To achieve an advantage to stimuli presented at the selected location, an observer can attend to a particular region of space, commonly referred to as spatial attention (Moran and Desimone, 1985). However, humans often know more about the defining features of objects (e.g. "the blue car will pass me from behind") than precise spatial locations. Consequently, to achieve an advantage for stimuli

with known features (in the above example a color and a direction of motion), independent of their spatial location, an observer will attend to a particular visual feature. This second type of selective attention is commonly referred to as feature-based attention (Treue and Maunsell, 1996; Treue and Martinez Trujillo, 1999; Martinez-Trujillo and Treue, 2004).

Despite its recognized behavioral importance, the neural basis of feature-based attention has only started to be understood (for an early example see Motter, 1994). Notably, in monkeys, feature-based attention amplifies the response of neurons when attention is directed to the neuron's preferred feature and suppresses responses when attention is directed to the neuron's nonpreferred feature (Treue and Martinez Trujillo, 1999; Martinez-Trujillo and Treue, 2004; Maunsell and Treue, 2006). This “feature-similarity gain” mechanism modulates the firing rate of neurons tuned to an attended feature when the neuron receptive field is inside the current location of spatial attention and also when the neuron is driven by a stimulus outside the focus of spatial attention (Treue and Martinez Trujillo, 1999; Saenz et al., 2002; Martinez-Trujillo and Treue, 2004). Thus, it seems that feature based attention is largely independent of spatial attention.

In humans, the study of feature-based attentional processing in early visual cortices has previously been thought inaccessible to non-invasive imaging techniques such as fMRI due to methodological restrictions, particularly the relatively coarse spatial resolution and inability to record from single neurons. However, the

emergence of multivariate pattern decoding (MVPD, see chapter I) rapidly changed this view and the neural representation of basic visual features in humans is now an active area of research (for review see Haynes and Rees, 2006). Moving from the neural representation of basic visual features to feature-based attention enabled Kamitani and Tong to find feature specific neural patterns in the human brain according to the attentional selection of one of two overlapping orientations (Kamitani and Tong, 2005b) and direction of motions (Kamitani and Tong, 2006). Building upon these results, Serences and Boynton (2007) demonstrated specific patterns of fMRI signals associated with feature based attentional selection of one of two overlapping motion directions (Serences and Boynton, 2007a). In that specific study, Serences and Boynton (2007) further demonstrate the independence of feature based and spatial attention. They show spread of the influence of an attended feature from the neural representation of the attended stimuli to neural representations of unattended stimuli somewhere else in the visual field. This is a qualitative replication of earlier findings from monkey single cell recordings but now in humans (Treue and Martinez Trujillo, 1999; Bichot et al., 2005). Additionally, their experimental design also allowed restriction of stimulus evoked activity to one hemisphere only by showing their stimulus only on one side of the screen. With this paradigm they show that the underlying neural representation of the unstimulated hemifield shows a measurable modulation according the (feature-based) attentional selection (Serences and Boynton, 2007a).

However, a closer look at the apparent distinction of feature-based and spatial attention by Serences and Boynton (2007) reveals some important shortcomings: First, the authors chose to study direction of motion, a feature known to be represented in the human cortex in two distinct subregions: MT and MST (Dukelow et al., 2001; Huk et al., 2002). While area MT responds mostly to contralateral motion-stimulation, MST responds mainly to ipsilateral motion stimuli. Thus, the distinction between MT and MST in principle offers an ideal setup to study hemifield specific stimuli. However, Serences and Boynton only defined a region they call hMT+, an ROI described as 'likely biased in favor of area MT' (see methods of Serences and Boynton, 2007a). Thus, the authors specifically do not exclude the possibility that results and conclusions from hMT+ might be driven by activity from the (spatially attended) stimulus represented in MST. Moreover, 'unattended' as well as 'unstimulated' activity was measured exclusively from a size and position matched patch of cortex in the ventral part of each hemisphere. However, if feature-based attentional influences were truly the explanation of their results, a modulation of visual cortices entirely unrelated to the stimulated area (i.e. a dorsal area) could potentially provide a more convincing demonstration. A further critical shortcoming is that early visual cortices V2 and V3 did not show any feature-specific modulations in their activity patterns when a stimulus was present, but, confusingly, only when stimulation was absent (see significant results in Serences and Boynton, 2007a Figure 3D and 3E). This is perplexing as feature based attention might be expected to act more strongly on an irrelevant stimulus than on no stimulation (Saenz et al., 2002). Finally, the authors completely fail to

demonstrate any significantly feature-specific modulations in striate cortex (V1).

Building on their results concering feature based attention, Serences and Boynton (2007b) claim to have found the representation of behavioral choice in human V5/MT (Serences and Boynton, 2007b). This is done by using an ambiguous motion stimulus that is equally likely to be perceived in two directions. The ability to decode the perceived direction of motion (as indicated by the participants) is interpreted as the ability to decode the behavioral choice of participants to perceive this (and not the other) direction of motion. However, Kaul and Bahrami (2008) point out that due to methodological shortcomings an alternate interpretation of the result might seem more likely: Instead of behavioral choice, Serences and Boynton (2007b) might have measured feature based attentional modulation in the very same hMT+ area that they showed similar modulations in the first paper (Serences and Boynton, 2007a). This explanation is also more consistent with parallel results from monkey-electrophysiology, also using ambiguous motion stimuli (Williams et al., 2003). Finally, explaining the results of Serences and Boynton (2007b) with the neural correlates of feature-based attention rather than behavioral choice resolves the apparent contradiction between previous results from monkey electrophysiology (Shadlen and Newsome, 2001; Williams et al., 2003). Rather than monkey MT, these previous studies find oculomotor areas (intraparietal sulcus and frontal eye field) show a high correspondence in their neural response of a motion based behavioral choice in monkeys – Serences and Boynton specifically

do not replicate this finding. As a consequence, it is conceivable that MVPD applied to BOLD activation induced by the ambiguous moving dot display by Serences and Boynton (2007b) might have decoded the attended direction of motion, replicating the earlier findings rather than indicating the neural correlates of a behavioral choice (Kaul and Bahrami, 2008). However importantly, these methodological shortcomings do not rule out the possibility that early sensory cortices indeed contain a neural substrate of an imminent behavioral choice, especially not if this choice involves a tuning of a sensory feature represented in these cortices. In the case of early visual cortices, basic visual features like orientation or color are particularly known to have a neural representation in early visual cortices (Hubel and Wiesel, 1962, 1968; Wang et al., 1998).

### VI.1.1 Objectives and hypothesis of this study

This chapter presents a study of feature selective attention and its dissociation from spatial attention in human early visual cortices. To get around shortcomings identified in a previous study (Serences and Boynton, 2007a), I chose to study feature-based attention to the basic visual feature orientation. My goal is to show the representation of feature selective attention in early visual cortices, including striate cortex, preferably without alternate explanations. As a second goal, I intentionally build my experimental paradigm such that an actual measure of the behavioral decision in early visual cortices is possible, again circumventing weaknesses identified in a previous study (Serences and Boynton, 2007b).

My hypotheses for this chapter are, firstly, to replicate earlier findings of orientation-selective attentional modulations of neural activity patterns in early visual cortices V1, V2 and V3 (Kamitani and Tong, 2005b). Secondly, to establish the independence of feature based attention from spatial attention by demonstrating orientation specific modulations of the neural activity patterns in entirely unstimulated and stimulus-position unrelated areas of V1, V2 and V3. Thirdly, to investigate the possibility that the behavioral decision of attending one of two orientations is present in relevant early visual cortices before any stimulus is presented.

## VI.2 Methods

### VI.2.1 Participants

Eight healthy participants (4 male, mean age 27 years) gave written informed consent to participate in the study, which was approved by the local ethics committee. All participants had normal or corrected to normal vision and were new to the task. One participant was excluded from analysis due to chance-performance in the behavioral task.

### VI.2.2 Stimulus

During the main fMRI experiment, participants performed a visual detection task in a circular patch in the lower left or right quadrant while maintaining fixation centrally at all times (Figure VII-1). The patch was itself was five degrees of visual angle in diameter and four degrees removed from the fixation point along a +/-45 degree angled line. Counterbalanced across participants it was placed either in the lower left or right quadrant and is henceforth referred to as the 'attended patch'. During the main task, full contrast, overlapping oriented lines at 45° and 135° (1.5 cycles per degree) were flashed on a gray background at 1.43 Hz (on for 650 ms, off for 50 ms) with a randomized spatial phase. For the duration of an entire main task block one of the two orientations would consist of black lines, the other one of white lines, resulting in a structured looking black and white plaid. Which orientation appeared in black or white was chosen at random. Each plaid always appeared on screen for 650ms, followed by 50ms of black screen followed by another 650ms of plaid at a random displacement (0-1.5 degrees displacement at random for each set of lines). Participants were required to monitor the occurrence of infrequent (14%) pre-specified targets associated with one of the two sets of oriented lines of their choosing, and to respond by a button-press to each detected target (6 targets per block). Targets consisted of one set of oriented lines decreasing 30% in contrast. This contrast dimming occurred in both sets of lines in a pseudo-randomized fashion: it was impossible for a target to be immediately followed by another target and it was impossible for targets in both lines to occur at the

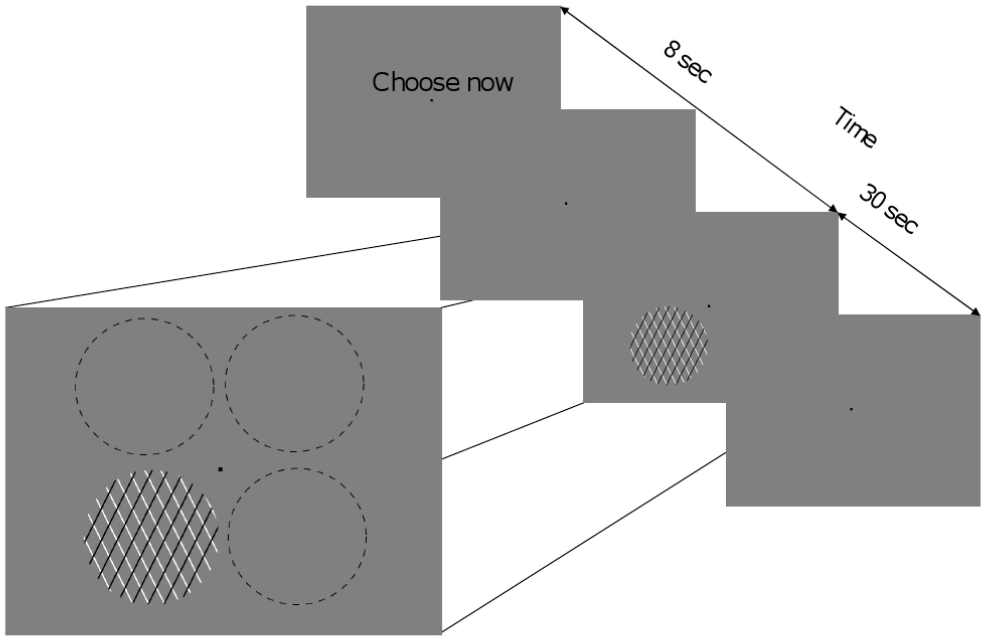



**Figure VI-1**
**Experimental stimuli. Each session contained four blocks, each consisting of an initial 8 second decision period, followed by a 30 seconds lasting experimental block, followed by a 20 second long rest period. As soon as the instructions "choose now" came on screen, participants were instructed to decide, at random, between attending a 45 or 135 tilted grating during the main task. During the 8 second delay, this decision had to be held in mind. The main task itself comprised overlapping oriented lines at 45° and 135° tilt, 1.5 cycles per degree, on a gray background flashed at 1.43 Hz (on for 650 ms, off for 50 ms) with a randomized spatial phase, displayed in either the left or right lower quadrant of the visual field, counterbalanced across subjects. Both gratings contained six targets: Participants indicated a slight dimming in the attended grating with a button press while continuously fixating on the central fixation dot. Eye movements were measured continuously during the experimental block. Indicated in the blown up representation in the lower left are the locations of the three size-matched patches localized during the patch localizer. Dotted lines were not on the screen during presentation.**

same time. Behavioural responses were collected with a standard MRI-button box, for a response to be counted correct it had to occur within a 900ms time window placed 200ms after stimulus onset.

In two separate sessions, participants viewed a reference stimulus to localize the retinotopic regions corresponding to the stimulated visual patch and equal size patched in all other quadrants. The "patch localizer" composed of high-contrast black and white flickering checkerboard patterns (10 Hz) presented in an each patch for 16.32 seconds, corresponding to 4 full volumes. Participant maintained fixation at all times, ensured by performing a counting task for a randomly disappearing fixation dot. Figure VII-1 illustrates the positions of these patches on screen.

Finally, in 2 additional sessions, standard retinotopic mapping localization procedures were performed to delineate visual areas on flattened cortical representations (see chapters III-V for details).

### VI.2.3 Procedure

Subjects lay supine in the scanner and viewed visual stimuli that were projected from an LCD projector (NEC LT158, refresh rate 60 Hz at 1024 X 768) onto a screen viewed via a mirror positioned within the MR head coil. Stimuli were presented using MATLAB (Mathworks Inc.) and COGENT 2000 (www.vislab.ucl.ac.uk/Cogent/index.html). Complete darkness was achieved in the scanning environment by manually masking the fMRI projection screen, head coil and internal bore with matt black card. This eliminated discernable non-retinotopic luminance cues, ensuring that the only source of visual stimulation during experimental sessions was the experimental stimulus. Eye position

was continually sampled for all participants at 60 Hz using an ASL 504 LRO infrared video-based MRI compatible eye tracker (Applied Science Laboratory, Bedford, MA).

All participants were scanned for a total of ten sessions, each containing four experimental blocks. Each block was preceded with the appearance of the instructions "Choose now" for two seconds, followed by a variable four to eight seconds delay interval with only a central fixation dot on screen. Participants were instructed to choose, at random, one of the two sets of oriented lines to attend to during the main task block to follow, then hold that decision in mind and not change it at any time during the delay period and task. The main task block lasted 30 seconds. Each block was then followed by a 20 second long rest period to give participants a short rest. Rest comprised only a fixation cross. Participants were instructed to fixate on the fixation cross for the duration of the entire 4 minute session including rest periods.

### VI.2.4 Scanning

A 3T Siemens Allegra system acquired T2*-weighted Blood Oxygenation Level Dependent (BOLD) contrast image volumes using a descending sequence every 4.08s. Each volume comprised 40 slices with a slice thickness of 1.5mm, positioned on a per subject basis to give coverage of the occipital lobe with an in-plane resolution of 1.5x1.5 mm. To maximize signal to noise in early visual cortex an occipital head coil was used. Each main

experimental session lasted 240seconds (59 volumes) split into 4 parts of 60 seconds each. To identify the boundaries of primary visual cortex (V1) and extra-striate retinotopic cortical areas V2 and V3, standard retinotopic mapping stimuli were presented in two sessions lasting 326 seconds each (80 volumes). Finally, the 'patch localizer' consisted of four repetitions of each quadrant plus rest for 5 volume each, summing to 408seconds (100 volumes) each session.

### VI.2.5 Analysis

*Behavioral analysis.* Subjects decided equally often to attend the 45 or 135 degree oriented lines. Within each block there were 6 targets. For the analysis only experimental blocks with more than 4 correct responses were considered. This led to an exclusion of a total of 64 out of 320 blocks (20%).

*Eye tracking Analysis.* The mean and standard deviation of eye movements were compared along the X and Y axis and the standard deviation of pupil size. This was done as to evaluate the possibility that a difference in eye movements between different conditions might be able to explain any differences subsequently found in brain activity. To this end, I first divided the X, Y eye position and pupil size values of each session by their mean (normalization) and then computed the mean and standard deviation. Any block in which the session standard deviation was over 2 standard deviations from overall mean of all participants was further excluded from

subsequent analysis (5/320 blocks, 1.5% of all blocks). Data from the remaining sessions, I evaluated for a difference between the two experimental conditions (45 or 135 degree oriented lines attended) with a two-sample t-test.

*Data preprocessing.* All data were preprocessed using Statistical Parametric Mapping software (SPM5, http://www.fil.ion.ucl.ac.uk/spm). The first 3 volumes of each fMRI session were discarded to allow for magnetic saturation effects. The remaining functional image volumes were realigned to the first image, and then the structural scan of each participant was co-registered to their functional data. Three single sessions displayed increased head movement and, consequently, they were also excluded from further analysis (12/320 blocks, 3.8% of all blocks). Functional data of the main experiment were not spatially smoothed, but data from the meridian mapping and 'patch localizer' sessions were spatially smoothed with a Gaussian kernel of 3mm full-width half-maximum.

*ROI localization*. FreeSurfer (http://surfer.nmr.mgh.harvard.edu) was used for segmentation and cortical flattening using each participant's specific structural image (previously coregistered to all functional data). Standard meridian mapping procedures were employed to identify the borders of the visual areas V1, V2 and V3 in the occipital cortex (Sereno et al., 1995). This was possible in all but one participant where meridian mapping data were insufficient to distinguish the borders of V3; consequently I left V3 undefined for this participant. To extract activity from these ROIs, I created

mask volumes for each ROI, these mask volumes should contain gray matter voxels only.

For the patch-specific ROIs, I defined subregions within V1, V2 and V3 that displayed increased activity for the patch localizer stimulus. Again, I then created mask volumes for each Patch-ROI.

*Timecourse analysis.* To compute the percent signal changes in the time course analysis, I used the mean raw activation of the realigned timecourse correcting only for slow signal drift typical in fMRI scanning by high-pass filtering (cut-off – 128s). Each session was then scaled by its own mean activity during the entire session. This linear transformation was repeated for each ROI and each participant separately. The resulting time series were sorted and averaged in each ROI according the two experimental conditions: 45 degree oriented lines attended vs. 135 degree oriented lines attended.

*Multivariate pattern classification.* Multivariate pattern analysis used in this chapter was identical to previous chapters and is thus only briefly described (for detailed description see chapter II and III-VI,): Unsmoothed, realigned fMRI data from the experimental runs were adjusted for the lag in hemodynamic response function by shifting all block-onset timings by 2 volumes (8 seconds). To determine classification accuracy, only classification with unseen and independent-test data was considered. Pattern classification was performed using a sparse logistic regression (SLR) algorithm (Yamashita et al., 2008). Significant difference from chance (2

categories = 50% chance) was tested for with a Student's one-sample t-test, applying Bonferroni correction for multiple comparisons across all ROIs examined. MVPD was repeated once for the main experiment and once for the 8 second period preceding the main experiment in which participants were asked to hold a decision in mind. Indicated in Figure VII-2 are the timing of the onsets of volumes during scanning and which volumes contributed to which result (dark gray for decision period and light gray for experimental block period). MVPD of subjects decision was only performed on 7 subjects as one subject applied a predestined logic to her attended oriented lines (always alternating).

Evaluating the probability that the classification was driven by over-fitting of arbitrary patterns of spatial correlations in the data, we carried out a shuffle-control test (Mur et al., 2009). If the assumption that classification is driven by chance were true, similar results should be obtained if labels indicating the experimental condition for each example vector were shuffled randomly. To test this, we ran a separate analysis where labels of the test examples were re-shuffled for each round of the cross-validation procedure. The resulting distribution of classification accuracy characterized the expected distribution of accuracy under null hypothesis.

## VI.3 Results

### VI.3.1 Behavioral results

In both attention conditions; that is, attention to one of the two sets of oriented lines as chosen at random by the participant, performance was similar in the subset of sessions and blocks within those sessions examined (see Methods). Both conditions had a roughly equal number of blocks: Participants chose to attend the 45 degree oriented lines 121 times, and the 135 degree oriented lines 124 times. Reaction times (RT) across participants of blocks attending to 45 vs. 135 degree oriented lines were compared with a paired, two-tailed t-test. [Reaction times 45 degree: 634ms; 135 degree: 629ms; $t(245) = 0.47$, $P = 0.64$]. Also, hit rates of correctly detected targets in both conditions were not significantly different [Correct responses 45 degree: 92.7%; 135 degree: 93.3%; $t(245) = -0.55$, $P = 0.59$].

### VI.3.2 Eye tracking results

For all participants, eye tracking data were sampled continuously throughout experimental blocks. In depth inspection of per-block mean and standard deviation of eye position and pupil size showed that most participants held fixation in most blocks. However, all blocks that displayed more than 2 standard deviations from the mean of all blocks in X or Y coordinates data were excluded (see methods). For all blocks considered three independent t-tests for X,

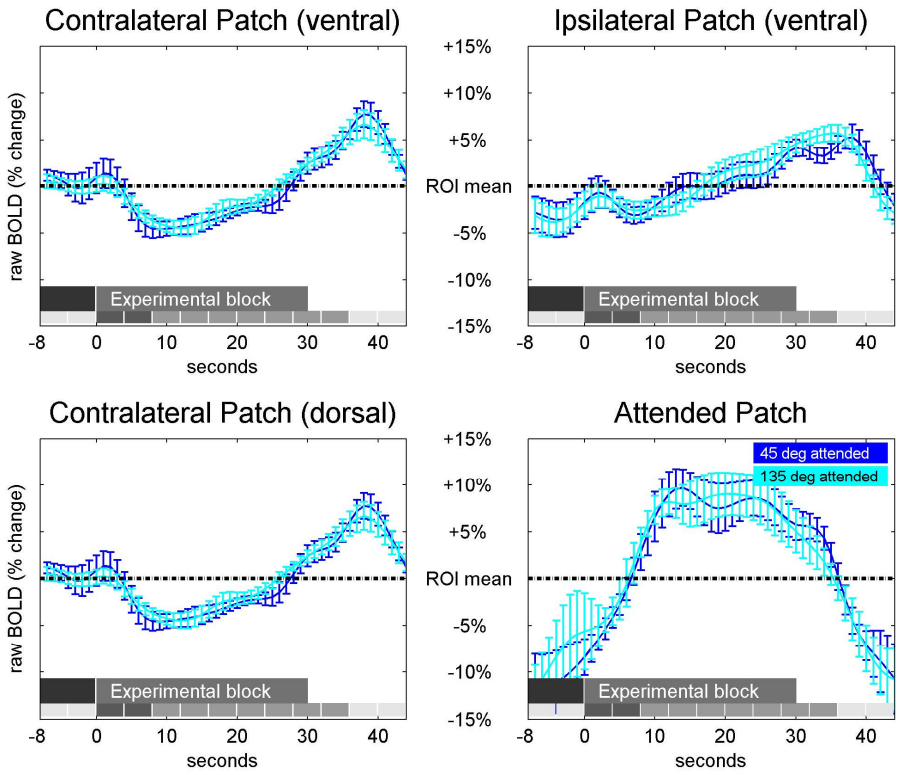



**Figure VI-2**
**Timecourses of the mean BOLD signal changes as a function of hemisphere and dorsal/ ventral areas (8 participants). Activity from each patch representation in V1, V2 and V3 averaged. The Y axis illustrates the percent signal change in comparison to activity to the ROI-session mean. The X axis depicts time in seconds aligned to each block onset. Each stimulus block was 30 seconds long, indicated by the gray bar labeled 'Experimental block'. Preceding each experimental block was an 8 second long delay period in which participants were required to hold in mind a desicion between attending to the 45 or 135 degree grating (indicated by the dark gray bar). Also indicated below the bars are image volumes as actually recorded during the experiment and as used for MVPD (delayed by 8 seconds to adjust for hemodynamic lag). The attended (stimulated) patch showed a clear increase of BOLD activity shifted by the hemodynamic delay. However no notable changes occur during neither desicion nor attention periods in any other hemisphere Patches.**

Y positions and pupil size did not show any significant differences between eye movements obtained form the two experimental conditions [movement along X axis: t(70) = -0.77, p = 0.45; movement along Y axis: t(70) = -0.69, p = -.49; Pupil size: t(70) = 0.23, p = 0.8].

### VI.3.3 Timecourse observation

Figure VII–2 shows the timecourse of raw BOLD signal changes for each participant and in each ROI in units of difference to the ROI-session mean. Time zero marks the onset of the 30 second long blocks indicated by the gray bar, preceded by the 8 second decision period in which participants were required to hold in mind a decision between subsequently attending to the 45 or 135 degree grating. The attended (stimulated) patch showed a clear increase of BOLD activity during experimental blocks, shifted by the hemodynamic delay. However, no changes occurred during either decision or attention periods in any other visual quadrant patches. Only a minimal offset increase at the end of the task was visible. Other than the expected outcome of increased visual activity at the location of the stimulus during the stimulated time, comparing the timecourses for blocks where 45 or 135 degree oriented lines were attended revealed no obvious differences. As a consequence, any further univariate analysis between the two conditions was unlikely to reveal anything further and is thus omitted.

### VI.3.4 Multivariate classification result (main task)

All classification results were obtained with SLR- MVPD. Figure VII-3 left illustrates the decoding result for the mask image of the attended patch from V1, V2 and V3: each representation contained sufficient feature specific information to predict significantly above

chance which of two oriented lines (45 or 135 degree tilt) was attended. [One tailed t-test against 50% chance, Bonferroni corrected, V1: mean = 59.3%, $t(7) = 3.9$, $P < 0.01$; V2: 60.8%, $t(7) = 3.8$, $P = 0.01$; V3: 60.5%, $t(6) = 4.5$, $P < 0.01$,].

I followed this procedure to obtain decoding results from every size-matched patch representation in V1, V2 and V3, both hemispheres, dorsal and ventral. BOLD activity in all but one patch representations allowed prediction significantly above chance with relatively little variation in the overall results [one tailed t-test against 50%: ipsilateral, ventral: V1: mean = 58.4%, $t(7) = 2.3$, $P = 0.03$, V2: 58.1%, $t(7) = 4.6$, $P < 0.01$; V3: 57.4%, $t(6) = 4.5$, $P < 0.01$; contralateral, ventral: V1: 56.4%, $t(7) = 3.2$, $P < 0.01$; V2: 56%, $t(7) = 4.6$, $P < 0.01$; V3: 55.6%, $t(6) = 2.6$,, $P = 0.02$; contrallateral, dorsal: V1: 52.9%, $t(7) = 1.8$, $P = 0.06$; V2: 54.9%, $t(7) = 2.3$, $P = 0.03$; V3: 58%, $t(6) = 2.7$, $P = 0.02$]. A one-way analysis of variance (ANOVA) also showed no significant differences between results from any of the 9 unattended patch representations [$F(8) = 0.63$, $p = 0.74$]. As a consequence, I therefore averaged across unattended patches per visual area for all further analyses.

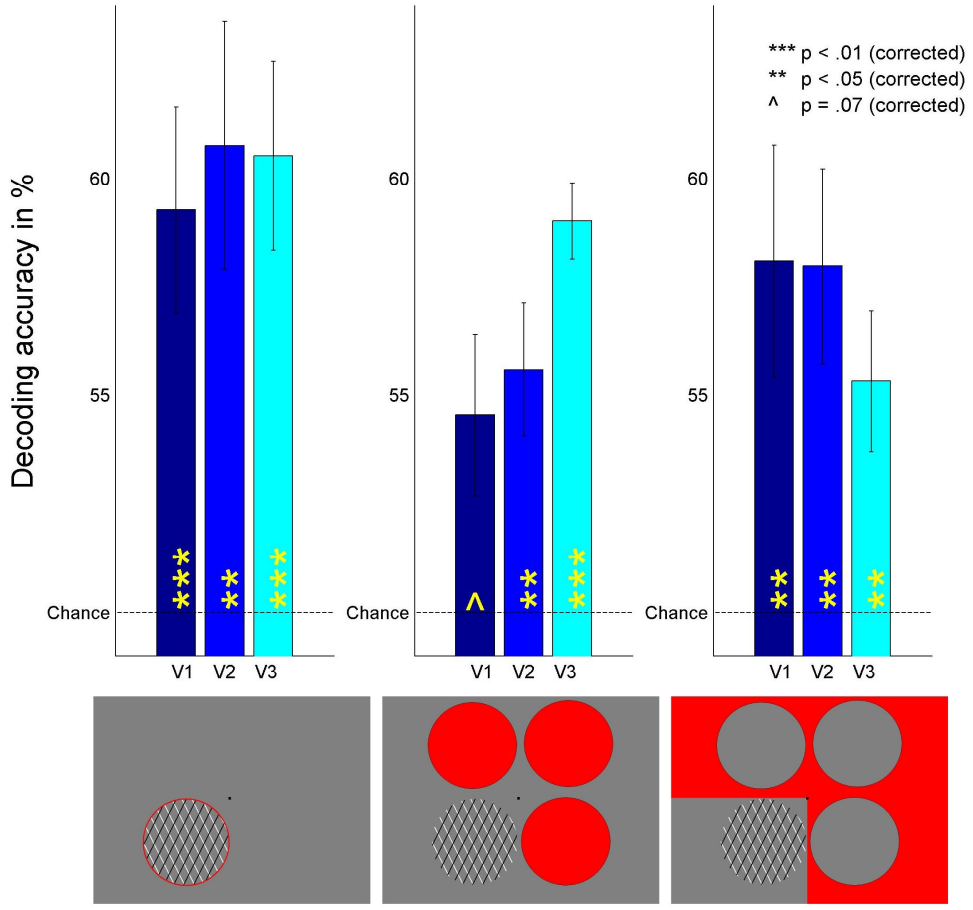



**Figure VI-3**
**Decoding accuracy of main experiment (mean across 8 participants) obtained with Sparse Logistic Regression (SLR). Different colors indicate decoding performance form different sub-areas in V1, V2 or V3. Yellow stars indicate results significantly different from chance: *** = p < .01, ** = p < .05 (one tailed t-tests), errorbars denote SEM across participants. Left: BOLD-activity patterns from the stimulated patch predicts significantly above chance which of the two orientations was attended in V1, V2 and V3. Middle: averaging across unattended patches, classification accuracy is still significantly above chance in V2 and V3. Right: Classification accuracy from the remainder of V1, V2 and V3 after subtracting the entire visual quadrant representation which contained the attended stimulus and the representations of size matched patches from all other visual quadrants: Each such reduced representation still contains enough feature-based information to decode attended grating orientation significantly above chance. All results are Bonferroni corrected for multiple comparisons.**

Figure VII-3 middle shows the overall result of BOLD decoding from unstimulated and unattended patches. Despite nearly no change in the BOLD signal when averaged across each ROI (see figure VII-2)

the spatial pattern of representations contained in these patches still allowed accurate prediction of attended orientation [One tailed t-test against 50% chance, Bonferroni corrected, V1: mean = 54.6%, $t(7) = 2.5$, $P = 0.066$; V2: 55.6%, $t(7) = 3.7$, $P = 0.01$; V3: 59%, $t(6) = 9.6$, $P < 0.01$].

Finally, I excluded voxels contained within the mask images from all unattended patches and the entire representation of the visual quadrant that contained the stimulus from V1, V2 and V3. Figure VII-3 right illustrates that, even when excluding all of these representations, the remaining parts of V1, V2 and V3 still contain enough pattern information in the BOLD signal to allow decoding significantly above chance [One tailed t-test against 50% chance, Bonferroni corrected, V1: mean = 58.1%, $t(7) = 3$, $P = 0.03$; V2: 58%, $t(7) = 3.6$, $P = 0.01$; V3: mean = 55.3%, $t(6) = 3.1$, $P = 0.03$].

To check the possibility that classification was driven by stimulus unrelated spatial correlations in the data (independent of the attended orientation), I carried out a shuffle-control test (Mur et al., 2009). To do this, I repeated the classification from the attended patches, but this time with shuffled labels for the test examples. None of the predictions are significantly different from chance. This control analysis strengthened the main findings, as it reconfirmed the validity of the result and the independence of the data used to obtain them [Bonferroni corrected results, V1: mean = 51.9%, $t(7) = 1.1$, $P = 0.49$; V2: 48.4%, $t(7) = -1$, $P = 2.48$; V3: 49.2%, $t(6) = -0.79$, $P = 2.31$]. I also repeated the same shuffle control

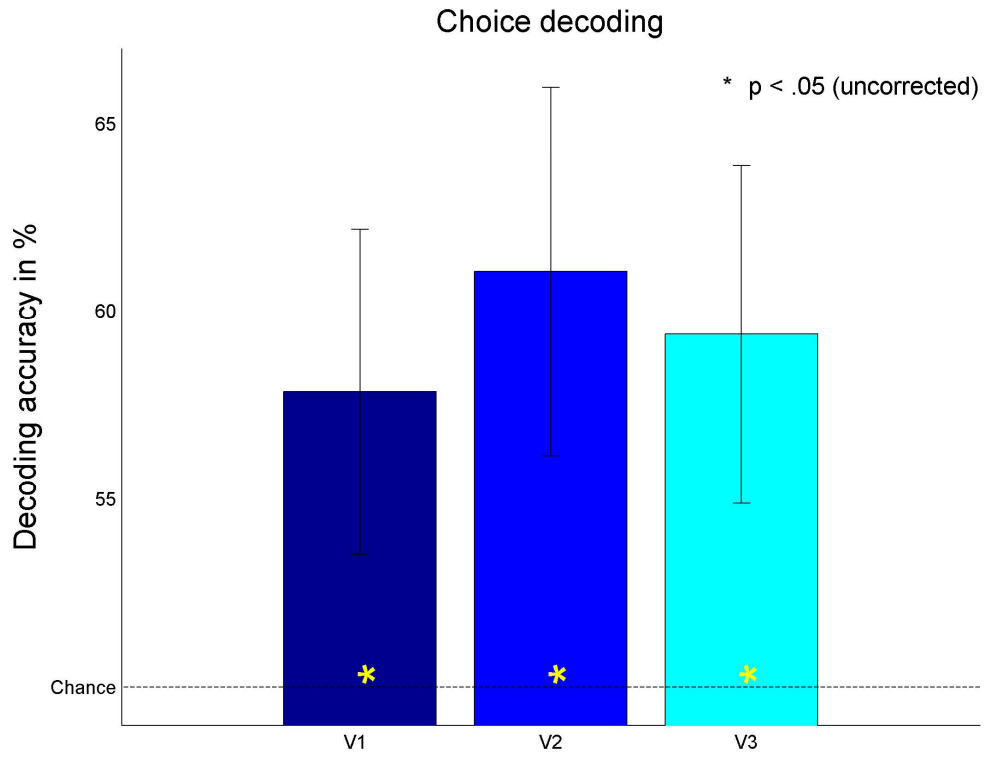



**Figure VI-4**
**Decoding accuracy of decision period preceding the experiment (mean across 7 participants) obtained with Sparse Logistic Regression (SLR). Different colors indicate decoding performance from V1, V2 and V3. Yellow stars indicate results significantly different from chance: * = p < 0.05 (one tailed t-tests), errorbars denote SEM across participants. BOLD-activity patterns from V1, V2 and V3 allow significant above chance decoding of which of two orientations a participant will attended to subsequently. Thus, all these early visual areas contain a representation of the decision subjects made. This decision is decodable prior to stimulus onset, is thus independent of stimulus evoked activity. Since subjects only indicated their decision only after each experimental block and each decision occurred at random, this prediction result is also independent of behavioral influences.**

analysis also for all other ROIs and obtained highly similar, non-significant results.

### VI.3.5 Multivariate classification result (decision period)

I repeated MVPD, this time focusing only on the 8 second delay period prior to stimulus onset. At beginning of this delay period

subjects were instructed to decide between subsequently attending 45 or 135 degree tilted oriented lines at random and hold their decision in mind during the delay. This 8 second delay period resulted in 2 full imaging volumes from which I attempted to decode the decision subjects were holding in mind. Note that no stimulus was visible during this period. Figure VII-4 illustrates the result of this decision-decoding: successful above chance prediction in all early visual areas [one tailed t-test against 50%, V1: mean = 57.9%, t(6) = 2, P = 0.048; V2: 61.1%, t(6) = 2.4, P = 0.03; V3: 59.4%, t(6) = 2.3, P = 0.03]. Possibly due to the low number of example vectors, results showed an increased overall variance compared to orientation-attention decoding (see figure VII-4).

## VI.4 Discussion

Here participants were presented with an initial 8 second decision period, followed by a 30s experimental block. During the initial decision period, participants were instructed to decide, at random, between attending a 45 or 135 oriented lines during the main task and hold their decision in mind. Then, during the main task participants detected a slight dimming in the attended oriented lines – which could be either 45 degree or 135 degree tilted and appeared in a circular region in one visual quadrant only while all other visual quadrants only displayed a medium gray background. Using MVPD, I was able to demonstrate that BOLD activity patterns within the representation of the stimulated patch contained spatial activity patterns that could distinguish the two attentional

conditions significantly better than chance. This replicates a previous finding that feature-based attention had a modulatory effect on underlying neural activity in early visual cortices (Kamitani and Tong, 2005b), but now for stimuli contained entirely within one quadrant of the visual field.

Further, I found significantly different activity patterns for the two feature-based attentional conditions across other regions of early visual areas V1, V2 and V3. This was seen for size-matched regions in other visual quadrants as well as other areas of early visual cortices even when completely excluding the stimulated quadrant plus size matched representations in all other visual quadrants (Figure VII-3). Attending to a particular specific orientation (feature) modulated orientation-selective (feature-selective) units across the entire visual field. This modulation was measured with fMRI in multiple regions of early visual cortex: V1, V2 and V3.

Finally, investigating a delay period preceding the main task, during which participants had to hold a decision in mind, I demonstrated that BOLD activity from early visual cortices during that delay period were sufficient to decode which orientation a participant will subsequently attend. Thus, these activity patterns contained a representation of the behavioral decision that a participant had made and was about to execute. This result is, to my knowledge, the first time that a behavioral decision was decoded from early visual areas without them having received any visual information.

Visual attention research has mainly focused on spatial attention, i.e., the attentional selection based on the current region of interest in the visual field. However, attention can be allocated not only to a particular region of space, but also to a visual feature, such as a particular color, orientation or direction of motion. Recordings from monkey cortex have demonstrated neural correlates of this type of attention in both the ventral and dorsal visual pathway (Maunsell and Treue, 2006). Notably, it seems that feature based attention in monkeys is largely independent of spatial attention as it occurs both inside the current location of spatial attention and also when a neuron is driven by a stimulus outside the focus of spatial attention (Treue and Martinez Trujillo, 1999; Saenz et al., 2002; Martinez-Trujillo and Treue, 2004).

In humans, Serences & Boynton (2007) proposed that allocating attention to one or the other of the two superimposed motion-surfaces, presented in one quadrant of the visual field, produced a differential pattern of activation across the neural representation of this patch and a size matched, unattended, patch in the comparable quadrant in the opposite hemisphere. This was possible for BOLD representations encoded in the size matched, unattended, patch in the opposite hemisphere both when a stimulus was present and absent. However, this study has numerous shortcomings: Firstly, it seem perplexing that the authors have not corrected for multiple comparisons across all their t-tests when comparing decoding results against chance predictions. This is while reporting 8 out of 10 "significant" feature-based-attention results on t-tests with $p < 0.05$ (Serences and Boynton, 2007a). Added to that, the authors

did not demonstrate significant feature-based attention effects in V1, V2 or V3 when a stimulus was present (see significant results in Serences and Boynton, 2007a Figure 3D). This is surprising as feature-specific enhancement of an irrelevant stimulus would be expected to modulate the perception of this irrelevant stimulus in favor of the attended feature. This phenomenon has been described as perceptual "tagging" (Saenz et al., 2002) and should have been measurable in the study by Serences and Boynton (2007a). However, the authors do report feature based attention effects in V2 and V3 in the absence of stimulation (see significant results in Serences and Boynton, 2007a Figure 3E). The combination of feature-based attention being absent when an irrelevant stimulus containing the attended features is present, but being present when a stimulus is absent is difficult to interpret. Thus, until now it had remained elusive whether early visual cortices V1 to V3 actually show feature specific modulations. Furthermore, the authors point out that their area hMT+ might contain voxels of area MST which is known to represent ipsilateral (i.e. attended) stimulus representations. Thus, this offers the possibility that results in Serences and Boynton (2007a) were actually driven by MST voxels. Therefore, in conclusion, it remains elusive whether early visual cortices V1 to V3 and area V5/MT actually contain feature based attention modulations that can be measured with fMRI.

In this experiment, I demonstrated that attending to one of two overlapping orientations elicits differential BOLD activation patterns at the spatially attended location, at size-matched locations in other parts of the visual field and, in fact, at all other locations of the

visual field as well. This shows that attending to a particular orientation in one location of the visual field specifically modulates orientation-selective units across the visual field. The modulating influence of feature-based attention was present in the attended area (figure VII-3 left) and multiple unattended areas where the stimulus was never shown (figure VII-3 middle and right). The most likely explanation seems that the underlying neuronal population of early visual cortices, irrelevant of retinotopic location, is modulated by the attended feature in the absence of stimulation. This result nicely parallels similar findings in the realm of spatial attention where response modulations have been observed in the absence of visual stimuli within the receptive field (Luck et al., 1997; Womelsdorf et al., 2006). As the stimulus was actually never shown anywhere else but in a small patch in one visual quadrant, this rules out the possibility that the measured feature-specific enhancement reflects some aspect of a irrelevant distractor (perceptual tagging) (Saenz et al., 2002). Rather, this result strengthens the hypothesis that a truly spatial-stimulus-independent feature-specific modulation was measured by MVPD fMRI.

Recently it was argued that the representation of perceptual choice between two opposite motion directions is contained in the pattern of BOLD activity of hMT+ (Serences and Boynton, 2007b). This result was based on using an ambiguous motion stimulus that was equally likely to be perceived in two directions. Participants were instructed that one direction was dominant and had to indicate which motion direction they perceived. Since there actually was no dominant motion direction present, the decoding of the reported

direction was interpreted as the ability to decode a behavioral choice of participants to perceive one (and not the other) direction of motion. However, subsequently this result was discussed as questionable (Kaul and Bahrami, 2008) as the experimental paradigm utilized did not control for the possibility that participants simply attended more to one direction (particularly because they were told it was present). In this case, the result by Serences and Boynton (2007b) would be a replication as it would have measured feature based attentional modulation in the very same hMT+ area that the authors showed similar modulations in an earlier study (Serences and Boynton, 2007a). Kaul and Bahrami (2008) further determine that this alternate explanation would also eradicate the perceived differences between the results by Serences and Boynton (2007b) and monkey studies using similar experimental procedures (Shadlen and Newsome, 2001; Williams et al., 2003). Thus, it remains elusive whether area V5/MT is modulated by a behavioral decision. The same question has never even been addressed for early visual cortices V1 to V3.

Here, I demonstrated that the behavioral decision of participants was represented in the BOLD activation patterns in early visual cortices V1, V2 and V3 (figure VII-4). The decision period preceded the stimulus presentation and subjects chose freely which orientation to attend. Thus, I controlled for both stimulus and behavioral independence. This setup overcame critical issues pointed out in previous studies (Serences and Boynton, 2007b; Kaul and Bahrami, 2008). As a result, I demonstrated measuring a behavioral decision in an early visual area before stimulation began.

One possible speculation why the modulation of neuronal activity was measurable in early visual cortices simply has to do with the fact that the behavioral decision in question was between two orientations. Considering that orientation is represented in a columnar fashion across early visual areas (Hubel and Wiesel, 1962, 1968; Wang et al., 1998) it seems to follow logically that these areas must be somehow involved. While this study noticeably demonstrates this involvement, further research is necessary to determine the exact role of early visual cortex modulation due to a behavioral decision. For example early visual areas might play an active role in the determination of a behavioral choice, or, early visual areas receive feedback input, modulating their activity as soon as a decision has been formed in a higher area, or, alas, a combination between the two scenarios. A further possibility might be that a top down priming-bias towards one of the directions might potentially explain the results for the epoch preceding the main task..

## VI.5 Conclusion

Findings in this chapter help bridging the gap between previous electrophysiological recordings in monkeys and studies of human perception. By using functional imaging and MVPD, I demonstrate that feature-based attention is an important aspect in the early human visual cortex. The results further strengthens the assumed independence of feature-based attention from spatial attention in

humans, an assumption which was recently also proposed in monkeys (Treue and Martinez Trujillo, 1999; Martinez-Trujillo and Treue, 2004). Thus they illustrate that feature-based attention is a highly functional attention system selectively enhancing specific features in our visual environment resembling the currently attended set of features. This might potentially occur at the expense of information about less relevant aspects or features. However, this study also opens up new questions, for example how integration of the various types of attention identified so far (e.g., spatial and feature-based but also object or surface-based, etc.) might occur at the level of single neurons as well as at the level of BOLD activation patterns.

As a second result, this study presents evidence that a behavioral decision can be measured reliably in an early visual area. It thus shows an involvement of early visual cortices in behavioral decisions. Further research should investigate the exact role of this involvement.

I conclude that results of this study are in exceptional agreement between with result from comparable studies from single cell monkey electrophysiology. This demonstrates how both approaches inform can inform each other and how they can and should be combined to answer new questions.

# Chapter VII:

# Gender specific face processing in the human brain

## VII.1 Introduction

Faces are processed in a distributed network of brain areas (Ishai et al., 2005; Fox et al., 2008; Ishai, 2008). A "core system" has been proposed, comprising of three regions that mediate the analysis of invariant facial features: the fusiform gyrus (FG, also known as fusiform face area, FFA), the inferior occipital gyrus (IOG, also called the occipital face area) and the posterior superior temporal sulcus (STS) (Haxby et al., 2000; Ishai et al., 2005; Gobbini and Haxby, 2007). Additionally, the "extended system" includes regions that mediate the processing of changeable aspects of faces, such as mood and expression. The extended system includes limbic regions, such as the amygdala (AMG) and insula (Ishai et al., 2004; Ishai et al., 2005); the inferior frontal gyrus (IFG) (Ishai et al., 2005), and regions of the reward circuitry, especially the nucleus accumbens and medial orbitofrontal cortex (OFC) (Aharon et al., 2001; Ishai, 2007).

Different regions of the core and extended systems display greater brain activity when specific aspects of face processing are required

by task demands. For example, the FG/FFA and IOG are more active in processes that require the identification of individuals (Puce et al., 1995; Kanwisher et al., 1997; Ishai et al., 2000; Grill-Spector et al., 2004), while tasks emphasizing gaze direction and speech-related movements modulate STS (Puce et al., 1998; Calder and Nummenmaa, 2007). The amygdala and insula are implicated in processing faces with emotional context and facial expressions (Breiter et al., 1996; Vuilleumier et al., 2001; Ishai et al., 2004; Ishai et al., 2005), the IFG is activated during the processing of semantic aspects (Ishai et al., 2000; Leveroni et al., 2000) and finally the OFC is implicated in the processing of facial beauty, sexual relevance and reward value (Aharon et al., 2001; O'Doherty et al., 2003; Kranz and Ishai, 2006; Ishai, 2007).

Although identification of gender is a fundamental, automatic and effortless aspect of face perception, conventional fMRI data analyses have not to date identified any regions within the face network specialized for the discrimination of gender. Investigating fMRI adaptation to facial gender and race, one study in fact identified the strongest adaptation effects occurring outside the face network, in the cingulate gyrus, while the same subjects showed only weak adaptation effect in regions of the core face network (Ng et al., 2006). Another study, looking specifically at gender-related face processing, found no evidence for differential activation associated with facial gender across the core or extended face networks; rather, any gender-specific differences were modulated by the sexual preference of the participants (Kranz and Ishai, 2006). Taken

together, previous work has shown weak, inconsistent or otherwise qualified neural responses to the gender of faces.

Importantly, all previous work has used conventional univariate data analyses of functional MRI data that focus on single locations (voxels) within the brain. More recently, there has been much interest in findings that local spatial patterns of fMRI signals encode considerable more information about visual stimuli **REFS HERE**. I hypothesized that applying these new and potentially more sensitive analyses to the data previously collected by Kranz & Ishai (2006) might now show evidence for more subtle distributed representations of facial gender in the human brain encoded within brain areas associated with face processing. I therefore sought to identify brain regions that exhibited responses specific to the gender of faces using new multivariate pattern analyses for individual participants. The goals of these new analyses were to identify brain areas containing spatial activity patterns across presentation of faces of one gender versus faces of the opposite gender, such that they can be used to reliably identify facial gender in different scanning sessions. To that end, I compared activity patterns within the core and extended regions of the face network with three control regions: the cingulate gyrus and two non-cortical control regions. In addition, I also examined whether successful classiciation might arise from spatial patterns of fMRI signals in early visual cortex, suggesting a role for low-level visual properties rather than a high-level representation of gender. Finally, I carefully explored the new, multivariate analyses for any evidence that activation patterns associated with facial gender might vary as a

function of the participant's gender (man or woman) and/or sexual preference (hetero- or homosexual).

## VII.2 Methods

Of the methods of the study the data originated from (Kranz and Ishai, 2006) only those directly relevant to results and analysis in this Chapter are repeated.

### VII.2.1 Participants

Forty normal, right-handed participants (mean age 26 ± 3 years, 10 participants in each of four groups, homosexual and heterosexual men and women) with normal vision participated in the study. All participants gave informed written consent for the procedure in accordance with protocols approved by the University Hospital of Zurich. Participants were classified as heterosexuals or homosexuals based on their self-report in a modified version of the Sell questionnaire (Sell, 1996).

### VII.2.2 Stimulus

Stimuli were displayed with Presentation (http://www.neurobs.com) and projected with a magnetically shielded LCD video projector onto a translucent screen placed at the feet of the participant.

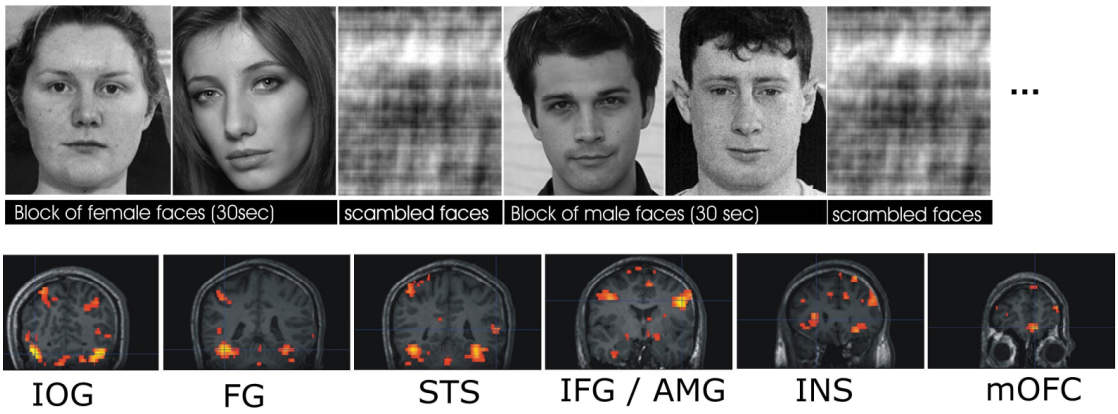



**Figure VII-1**
**Top: Examples of face stimuli. Each face was presented for 3 sec in alternating 30 second long blocks of either male or female faces, which alternated with scrambled faces. Participants viewed the faces (3 runs) or rated their attractiveness (5 runs). For additional details see (Kranz and Ishai, 2006). Bottom: Face responsive ROIs from one exemplar participant. Displayed is the overlay of the contrast "all faces vs. scrambled faces" onto the participants unflipped anatomical image (left hemisphere is on the left side). Sections show coronal slices, from posterior to anterior: IOG, FG, STS, IFG & AMG and INS with show bilateral activation, in contrast to the medially defined OFC.**

Participants viewed grayscale photographs of faces (three runs) and assessed facial attractiveness (five runs). In each run, 2 epochs of male and 2 epochs of female faces (both 30 s) alternated with epochs of phase-scrambled faces (21 s in viewing, 12 s in attractiveness rating). During an epoch each stimulus was presented for 3s, with no blank periods between the stimuli. In total, during the viewing condition, 60 male and 60 female unfamiliar, famous, and emotional faces were presented. During the assessment of facial attractiveness, 100 male and 100 female faces were presented. Faces were optimized for facial attractiveness ratings, thus they included facial hair, a variety of viewing angles and moderate size differences. The order of runs was randomized across participants. Importantly, in each session there was an equal amount of female and male faces. As participants were not

instructed to pay attention to the face gender, any gender processing was implicit during the two tasks. Figure 1 (top) shows examples of the stimuli used.

### VII.2.3 Procedure

Participants lay supine in a 3T Philips Intera whole-body MR scanner (Philips Medical Systems, Best, The Netherlands). Changes in the blood-oxygenation-level-dependent MRI signal were measured with the sensitivity-encoded gradient-echo echoplanar sequence (35 axial slices, TR = 3000 ms, TE = 35 ms, flip angle = 82°, field of view = 220 mm, acquisition matrix = 80 × 80, reconstructed voxel size = 1.72 × 1.72 × 4 mm, SENSE acceleration factor R = 2). High-resolution, spoiled gradient-recalled echo structural images were collected in the same session for all the subjects (180 axial slices, TR = 20 ms, TE = 2.3 ms, field of view = 220 mm, acquisition matrix = 224 × 224, reconstructed voxel size = 0.9 × 0.9 × 0.75 mm). These high-resolution anatomical images provided detailed anatomical information for the region-of-interest (ROI) analysis.

### VII.2.4 Analysis

*Data preprocessing.* The data were preprocessed using Statistical Parametric Mapping software (SPM5, http://www.fil.ion.ucl.ac.uk/spm). The first 5 volumes of each fMRI scan were discarded to allow for magnetic saturation effects. The

remaining functional images volumes were realigned to the first image, then the structural scan of each participant was co-registered to their functional data. Functional data were not spatially smoothed.

*Timecourse analysis, univariate analysis and ROI localization*

To compute the percent signal changes in the time course analysis, I used the mean raw activation of the realigned timecourse correcting only for slow signal drift typical in fMRI scanning by high-pass filtering (cut-off – 128s). Each session was then scaled by its own mean activity during all blocks of scrambled faces. This linear transformation was repeated for each ROI and each participant separately. The resulting time series were then sorted and averaged in each ROI according to male or female face blocks, thus resulting in a percent difference plot for faces of each gender in each region of interest.

I went on using SPM5 to perform a within-participant analysis, using a voxel-wise general linear model (GLM) that comprised 3 delayed boxcar waveforms representing the 3 experimental conditions: female faces, male faces and scrambled faces. During this analysis the fMRI time series were high-pass filtered (cut-off – 128s) and global changes in activity were removed by proportional scaling of each session. Then I computed the contrast of all faces versus all scrambled faces. Note that for each participant this contrast contains a balanced number of blocks containing female and male faces and is thus orthogonal to the experimental question of this study.

To identify the different areas of the core and extended systems of the face network, I overlaid the contrast of all faces vs. all scrambled faces at a FWE-corrected level of significance of P<.05 onto each individual participants structural image (Figure 1). After visually identifying all seven regions of the core and extended systems, the fusiform gyrus (FG, also known as the fusiform face area, FFA), the inferior occipital gyrus (IOG), the posterior superior temporal sulcus (STS), the amygdala (AMG), the insula (INS), the inferior frontal gyrus (IFG) and the medial orbitofrontal cortex (OFC), a sphere-shaped ROI was defined only for those regions that displayed activity over the threshold. The peak of the activation defined the centre of the ROI. To adjust for different sizes of brain structures I used a sphere with 10mm diameter for FG, IOG, insula, IFG and medial OFC and an 8mm diameter sphere for STS and the amygdala. Where face specific activity was not identified at a FEW-corrected threshold of P<.05, I did not define ROIs. All areas were collapsed across hemispheres resulting in one ROI per face region.

In addition to the face-responsive ROIs, four size matched, non-face-responsive control regions were anatomically defined. Control region one (CTR1) comprised the gray matter of the medial mid-cingulate closely matching the definition of the cingulate gyrus from Ng et al (2006). The second (CTR2) and third (CTR3) control regions represented non-cortical white matter of the corpus callosum and an area posterior to the pons, covering mostly parts of the fourth ventricle respectively. Finally, a fourth control region (OP) represented early visual cortex, comprising a region slightly

anterior of both occipital poles, medially, covering occipital sections of the calcarine sulcus of both hemispheres. I hypothesized that CTR1 would only show significant classification of facial gender under the alternate hypothesis that the cingulate gyrus contains gender specific information. CTR2 and CTR3 represented negative control regions that should not show any classification accuracy for facial gender and should therefore control for any non-specific artefacts. Finally, I included early visual cortex (OP) to investigate the possibility that any successful discrimination of facial gender in higher visual areas might instead be due to different low-level image characteristics represented in early visual cortex.

*Multivariate pattern classification.* Unsmoothed, realigned fMRI data from the 8 experimental runs were adjusted for the lag in hemodynamic response function by shifting all block-onset timings by 3 volumes. Then, data were transformed into "example vectors" for the classifier (Pereira et al., 2009). The ten volumes of each block resulted in ten example vectors, containing each voxel in the ROI. The resulting example vectors were concatenated to form a matrix whose rows and columns corresponded to category examples (male or female) and voxels in the ROI, respectively. Data from each voxel (i.e., each column of the matrix) were then z-normalized to have zero mean and unit variance. The resulting matrix, together with a label for each row indicating the stimulus condition was taken to the next stage.

To determine classification accuracy, only classification with unseen and independent test data was considered. Thus, test data sets in

different iterations were always independent of the training data sets used. I used a leave-one-out cross-validation method to evaluate the classification accuracy (Mur et al., 2009; Pereira et al., 2009). Because the data were obtained in 8 separate independent runs consisting of 20 volumes of each category, each test and training set consisted of 40 and 280 examples, respectively.

Pattern classification was performed using a sparse logistic regression (SLR) algorithm (Yamashita et al., 2008). SLR is a Bayesian extension of logistic regression that combines an innovative strategy for adaptive, yet unbiased voxel selection with the conventional linear discriminant analysis. Within every iteration of the cross-validation SLR carried out a number of *nested cross-validations* inside the training set: the training set was divided randomly in two sections of specified proportion; for a randomly selected subset of the voxels, the linear classifier was trained with one section of the data and tested with the other and the selected voxels were weighted proportional to the accuracy of this classification. This procedure was carried out 500 times and the voxels accumulated weights. At the end of the nested cross-validation, the assigned weight of each voxel was taken as a *relevance* factor indicating how informative the voxel was for classification. Voxels with the highest relevance were then selected for the actual classification. Supplementary table 2 illustrates the number of voxels chosen for each ROI. Importantly, this voxel selection algorithm depended entirely on the training set and was completely ignorant about and independent of the test set. The

training and test data from the selected voxels were then passed on to a conventional linear classifier (Yamashita et al., 2008).

Classification accuracies were averaged across the 8 cross-validations for each ROI in each observer data assignments. Thus, for each observer this procedure yielded exactly one prediction accuracy per ROI, i.e., 40 observations per ROI. I tested for a significant difference from chance (2 categories = 50% chance) with a Student's one-sample t-test, applying Bonferroni correction for multiple comparisons across ROIs examined. Where Bonferroni corrected p-values were greater than $p = 0.05$, they are simply reported as not significant (n.s.) expect when trending toward significance. I also repeated testing for a statistical significant difference with a two-sample t-test against a second null hypothesis of chance performance as defined by the distribution of classification accuracy in the three control ROIs, again Bonferroni correcting the result for multiple comparisons.

Finally, in order to evaluate the probability that the classification was driven by over-fitting of arbitrary patterns of spatial correlations in the data, I carried out a shuffle-control test (Mur et al., 2009). If the assumption that classification is driven by chance were true, similar results should be obtained if labels indicating the experimental condition for each example vector were shuffled randomly. To test this, I ran a separate analysis where labels of the test examples were re-shuffled for each round of the cross-validation procedure. The resulting distribution of classification

accuracy characterized the expected distribution of accuracy under null hypothesis.

## VII.3 Results

### VII.3.1 Behavioral results

The behavioral data confirmed that all participants, regardless of their gender or sexual preference, rated the attractiveness of male and female faces similarly (Kranz and Ishai, 2006).

### VII.3.2 Univariate results

I identified areas of the core and extended systems of the face network by overlaying the contrast of all faces vs. all scrambled faces at FWE = .05 onto each individual participants structural image. Figure 1 shows this contrast for one exemplar participant at different coronal sections. Figure 2 displays the mean time course for all ROIs, averaged across participants. Time zero marks the onset of the 30 second long blocks containing either male or female faces. Within all areas of the core and extended face network I found significant responses to face stimuli during the 30 second presentation, peaking about 6-8 seconds after block onset consistent with the timing of the haemodynamic response. In contrast, overall BOLD activity in the control regions (CTR1–3) did not show evoked responses that corresponded with either the

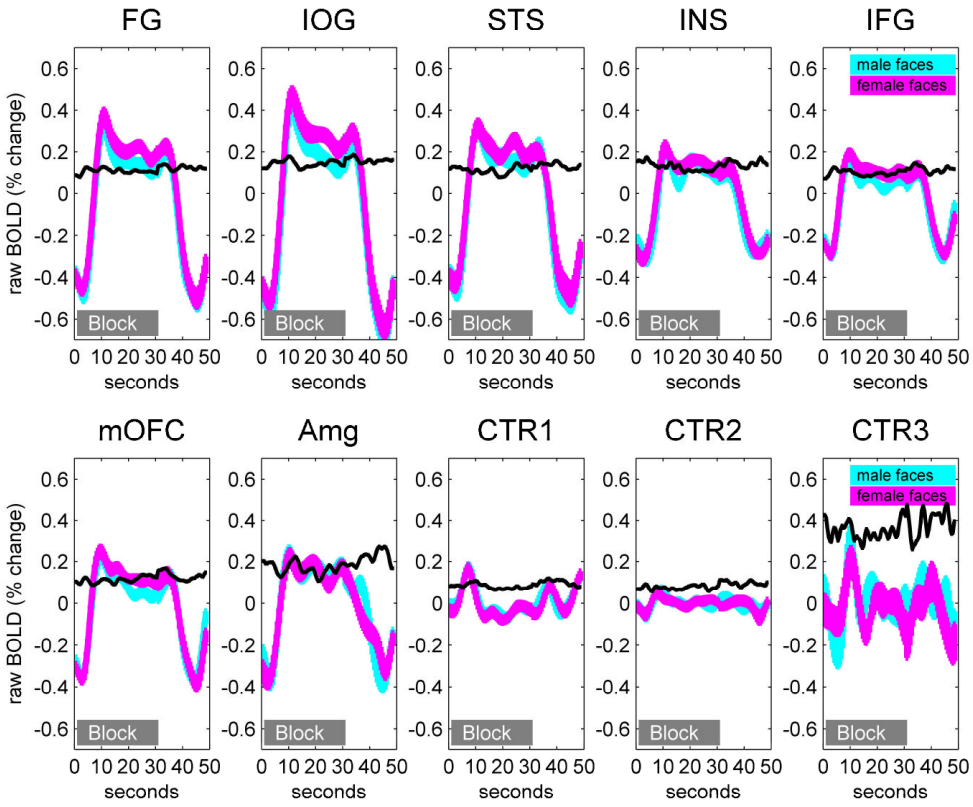



**Figure VII-2**
**Timecourses of the mean BOLD signal changes from all ROIs. Activity of blocks of female or male faces was averaged over all 40 participants. The Y axis illustrates the percent signal change in comparison to activity to scrambled faces. The X axis depicts time in seconds aligned to each block onset. Each stimulus block was 30 seconds long, indicated by the gray bar. All ROIs of the core and extended face processing network showed a clear increase of BOLD activity shifted by the hemodynamic delay. In comparison, control ROIs, CTR1 to 3, showed no clear activation pattern in relation to the stimulus blocks. Over all ROIs, activity differences between blocks consisting of male and female faces were inconsistent (cyan vs. magenta). Values in black depict the absolute difference between the two conditions. Across ROIs absolute differences were between 0.1 to 0.15%, irrespective of timepoint or brain region. Only CTR3, a non-brain ventricle, showed a dissimilar difference level between the two conditions, likely to be explained by its increased signal variance.**

onsets or durations of the blocks. BOLD signals in all ROIs were qualitatively very similar during the presentation of male and female faces (see Figure 2). To compare any individual differences in BOLD signal for blocks of female faces vs. blocks of male faces that might

be hidden in these group analyses, I computed the absolute difference between the two conditions for the entire time course and each ROI separately. In all ROIs this absolute per-participant difference between the two face gender conditions was very stable and was no greater than 0.1% BOLD signal difference at any time point (black line in Figure 2). A similar value of 0.1% absolute activity difference was found in the control regions within the brain, CTR1 and CTR2. Additionally, I computed a Pearson correlation for each pair of timecourses (male, female) for each ROI and subject, revealing high correlation between the timecourses of blocks containing male and female faces in FG, IOG, STS, INS, IFG and OFC (mean correlation RHO = 0.86) and low correlation between the same blocks in the amygdala and all control areas (mean correlation RHO = 49.5) [median RHO per ROI: FG = 0.93; IOG = 0.92; STS = 0.89; INS = 0.77; IFG = 0.8; OFC = 0.85; Amg = 0.65; CTR1 = 0.63; CTR2 = 0.42; CTR3 = 0.28]. Thus, the univariate analysis revealed highly correlated timecourses to blocks of male and female faces in FG, IOG, STS, INS, IFG and OFC but a lesser correlation in the amygdala and control regions.

### VII.3.3 Multivariate results

The left panel of Figure 3, left, shows the mean decoding accuracies (across all 40 participants) for all the ROIs identified as part of the core and extended face processing system, and all control areas. Gender could be identified significantly better than chance from BOLD signals in all three regions of the core network. In the core

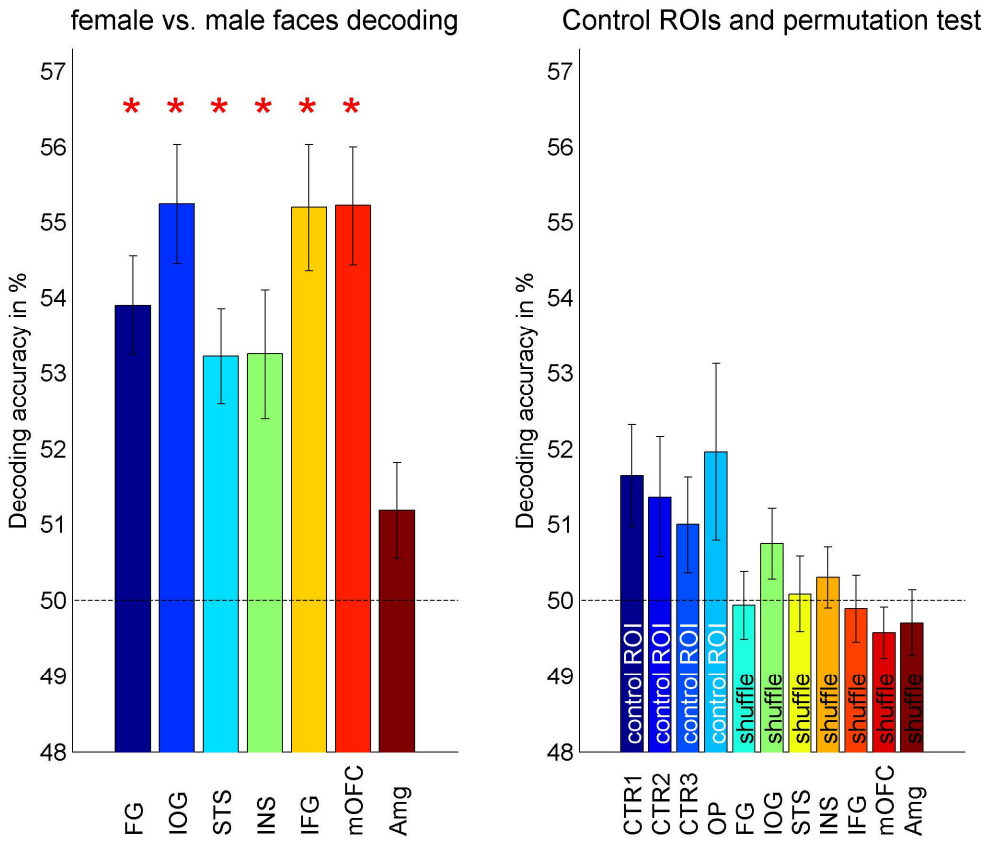



**Figure VII-3**
**Mean decoding performance for male vs. female faces in all ROIs. Left: Regions of the core (FG, IOG, STS) and extended (IFG, insula, OFC) show a significant difference from chance performance in predicting the gender of the presented faces. In the amygdala, however, no significant gender classification performance was observed. Right: Control regions consisted of sphere-shaped areas of non-face responsive gray matter, CTR1, white matter, CTR2, a ventricle, CTR3 and visually responsive cortex around the occipital pole, OP. No control-ROI shows any above chance prediction. Additionally, I performed a shuffle-control test (Mur et al., 2009) with randomly permuting labels for each set of test vectors in each cross-validation. Like the control ROIs, results of the permutation test are no significantly different from chance. Together, all control results confirm the distribution of classification accuracy expected under null hypothesis.**

network, the highest accuracy was obtained from the IOG followed by FG and STS [FG: 53.9%, $t(39) = 6$, $P < 0.001$; IOG: 55.2%, $t(39) = 6.6$, $P < 0.001$; STS: 53.2%, $t(31) = 4.6$, $P < 0.001$]. In addition, gender was decoded from BOLD signals significantly better than chance in three regions of the extended system: Activity

patterns from the inferior frontal gyrus (IFG), the insula and the orbitofrontal cortex (OFC) all permitted decoding of gender greater than chance [INS: 53.3%, $t(29) = 3.3$, $P < 0.02$; IFG: 55.2%, $t(34) = 5.8$, $P < 0.001$; mOFC: 55.2%, $t(25) = 5.4$, $P < 0.001$]. In contrast, BOLD signals from the amygdala (AMG) were not sufficient to allow for above-chance classification of gender information [Amg: 51.2%, $t(21) = 1.4$, n.s.].

These data were evaluated collapsing across hemispheres, but it is well recognized that the ventral visual pathway shows a degree of functional asymmetry in its responses to faces. To evaluate any possible differences in hemispheric classification accuracy, I therefore repeated the analysis in FG and IOG separately for each hemisphere. When considered separately, right and left FG and IOG successfully predicted facial gender at similar levels to that seen when (as above) analyzed together [IOG left: 54.1%, $t(37) = 5.5$, $P < 0.001$; IOG right: 54.5%, $t(39) = 5.5$, $P < 0.001$; FG right: 52%, $t(38) = 2.5$, $P = 0.05$; FG left: 53.9%, $t(38) = 6.7$, $P < 0.001$]. However, most importantly there were no significant differences comparing classification accuracies of left and right hemisphere IOG and FG across all subjects [paired ttest left vs. right: FG: $t(37) = -1.8$, n.s.; IOG: $t(37) = 0.25$, n.s.].

A number of control analyses were performed. Firstly, I attempted to classify gender information from three control areas that are not known or highly unlikely to contain gender information (Figure 3, right). Classification performance was poor in these control ROIs and did not differ significantly from chance [CTR1: 51.6%, $t(34) =$

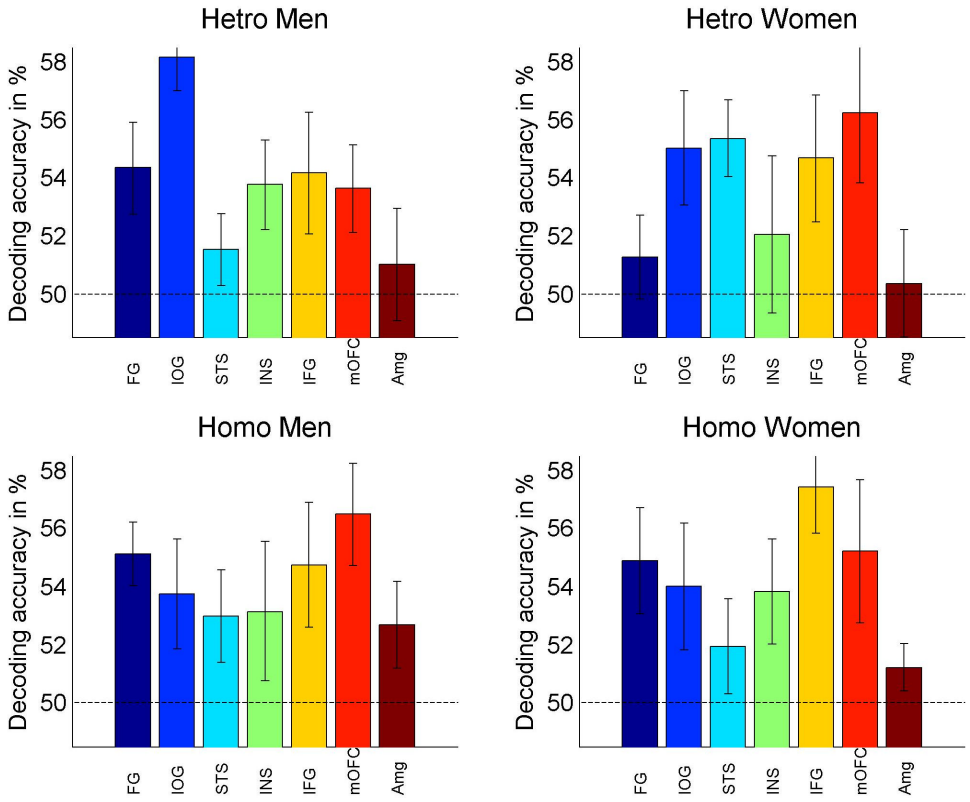



**Figure VII-4**
**Mean decoding performance as a function of the participant's sexual preference. The decoding profiles in each group were very similar to the mean averaged across of 40 participants shown in Figure 2. A separate ANOVA for each ROI reveals no significant differences between the groups.**

2.3, n.s.; CTR2: 51.4%, t(34) = 1.6, n.s.; CTR3: 51%, t(34) = 1.5, n.s.]. Thus our findings of successful classification of facial gender were specific to the face network.

I also evaluated whether facial gender could be predicted from patterns of activity in visual cortex. Again, classification performance was poor and not significantly different from chance [see Figure 3, right; OP: 52%, t(39) = 2.9, P = 0.07]. This indicates that successful classification accuracy in regions of the face network

was not driven by low-level properties of the faces represented in fMRI signals from early visual cortex.

To evaluate the probability that successful gender classification was driven by over-fitting of arbitrary patterns of spatial correlations in the data that were independent of the gender of the faces, I carried out a shuffle-control test (Mur et al., 2009). I repeated the classification from all face-responsive ROIs, but this time with shuffled labels for the test examples. The remaining bars in the left panel of Figure 3 depict the results of this shuffle-control analysis. Training the classifier using the same training sets but with shuffled labels for example vectors confirmed the distribution of classification accuracy expected under null hypothesis. This control analysis strengthened the main findings, as it reconfirmed the validity of the result and the independence of the data used to obtain them. [FG: 49.9%, $t(39) = -0.15$, n.s., IOG: 50.7%, $t(39) = 1.6$, n.s., STS: 49.6%, $t(31) = -1.1$, n.s., INS: 50.3%, $t(29) = 0.65$, n.s., IFG: 49.9%, $t(34) = -0.24$, n.s., mOFC: 49.7%, $t(25) = -0.55$, n.s., Amg: 50.1%, $t(21) = 0.12$, n.s.].

I tested whether decoding the gender of face stimuli depended on the gender or sexual preference of the participants. Figure 4 displays the decoding results for each group (namely, heterosexual men, homosexual men, heterosexual women, and homosexual women) separately. To test for differences between the groups, a one-way analysis of variance (ANOVA) was performed for each ROI. However, even when uncorrected for multiple comparisons, all 4 groups of participants showed very similar patterns of classification

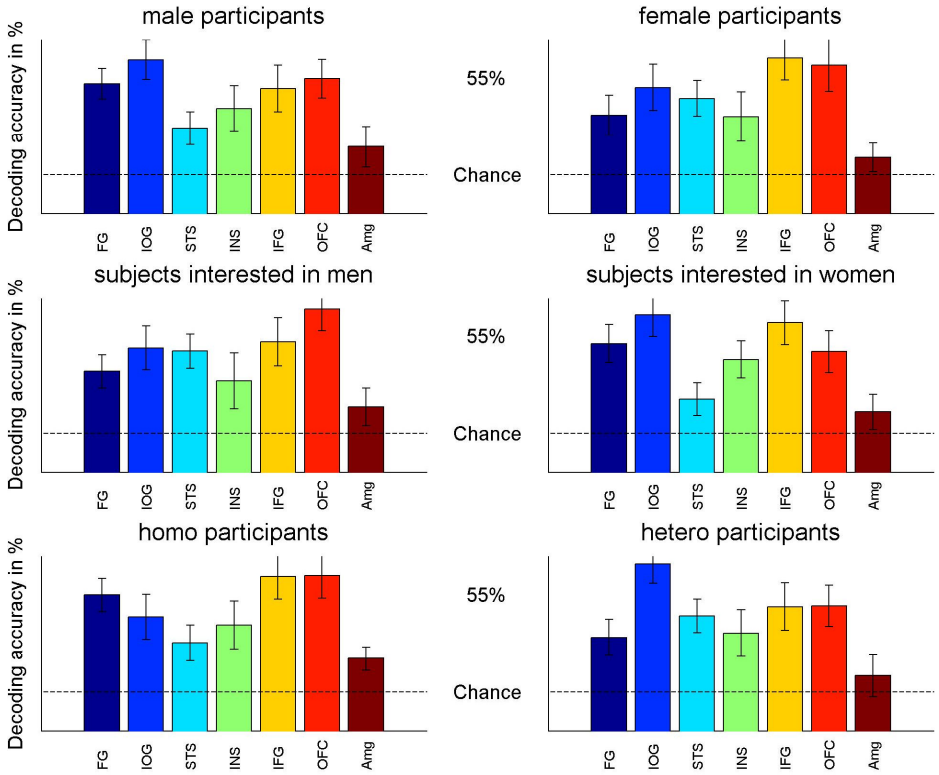



**Figure VII-5**
**Decoding performance in different sub-groups of the population of subjects. Top: male and female participants. Middle: participants interested in men and participants interested in women. Bottom: homo and hetero participants. Overall there were no large differences in decoding performance in any subgroup. A separate student t-test for each ROI and constellation of groups revealed no significant differences after correcting for multiple comparisons (see supplementary table 1).**

in all ROIs [group means in brackets (Hetero Men, Hetero Women, Homo Men and Homo Women), FG: n.s. (54.3, 54.9, 51.3, 55.1); IOG: n.s. (58.2, 54, 55, 53.8); STS: n.s. (51.5, 51.9, 55.4, 53); INS: n.s. (53.8, 53.8, 52, 53.1); IFG: n.s. (54.2, 57.4, 54.7, 54.7); mOFC: n.s. (53.6, 55.2, 56.3, 56.5); Amg: n.s. (51, 51.2, 50.4, 52.7)].

Further, I investigated whether the distribution of gender-information in the face network varied according to the gender of

participants (male vs. female participants), by gender-specific sexual preference (interested in men vs. women) or by sexual orientation (homo vs. hetero participants). I compared each pair of group-results with in each ROI with a two-sample t-test. Results are Bonferroni-corrected for multiple comparisons. Figure 5 shows difference graphs obtained by subtracting the results for each of the subgroup pairs. While the top graph depicts decoding accuracies between male and female participants, the middle graph shows decoding accuracies between participants interested in men versus participants interested in women. Finally the bottom graph displays accuracies between homo and hetero participants. There were no significant differences for any group constellations in any ROI (for detailed statistical test values see Supplementary table 1).

Finally, since the instructions during 3 out of 8 sessions in the original study were to simply view image of faces and not rate facial attractiveness, this might conceivably be problematic for the interpretation of the results as different task demands may affect face processing. In order to evaluate this possibility, I repeated the entire analysis for the 5 rating sessions only (excluding the passive viewing sessions). Supplementary Figure 1 displays the result of this analysis that replicated the main findings of this study. Thus I were unable to find any evidence that successful gender-specific classification was modulated by task performance, at least for passive viewing and rating tasks. This is consistent with gender classification being an implicit feature of both tasks.

## VII.4 Discussion

The goal of this study was to identify gender-specific patterns of activation in the human brain. Using data that were previously collected for a study identifying the neural correlates of facial attractiveness (Kranz and Ishai, 2006) across male and female faces, I mapped face responsive brain areas of the core (fusiform gyrus (FG), inferior occipital gyrus (IOG) and superior temporal sulcus (STS)) and extended (amygdala (AMG), insula (INS), inferior frontal gyrus (IFG) and orbitofrontal cortex (OFC)) face network (Ishai et al., 2005). I demonstrated that BOLD signals averaged over these ROIs in response to seeing either female or male faces was comparable in all of these areas and was not sufficient to distinguish facial gender. However, using multivariate pattern analysis, I showed that the local spatial pattern of BOLD signals from the FG, IOG, STS, INS, IFG and medial OFC all contained sufficient information to decode gender of observed faces significantly better than chance. I did not detect, however, significant gender classification performance in the amygdala, early visual cortex, and three other control regions. I further confirmed the specificity of our classification analyses using a shuffle-control test (Mur et al., 2009). Finally, we did not find any variability in our ability to classify facial gender according to the demographics of the study population (specifically, their gender and sexual orientation).

Taken together our findings indicate that, rather than localized to a single region and despite the fact that I could not find gender-specific increases in mean levels of brain activity in any single area

for the face-network, gender information is present in the pattern information within the core and most regions of the extended face-network. This result might implicitly herald the possibility that other aspects of faces might also be identifiable from different regions within the face network. The present study could not address this question as only facial gender was manipulated explicitly as a stimulus property. From our results I conclude that, within the face-network, gender-information is a highly distributed attribute of facial perception.

Perception of faces elicits activation within a distributed cortical network that includes the core and extended regions I focused on in this study (Haxby et al., 2000; Ishai et al., 2005). Additionally, one study reported fMRI adaptation effects for facial gender and race outside this network, specifically in the cingulate gyrus (Ng et al., 2006). I probed an anatomically defined cingulate gyrus ROI (CTR1), but did not find any significant gender specific activation with conventional univariate analysis, or any decoding results significantly different from chance with multivariate pattern analysis. However, Ng et al (2006) offered a possible reconciliation of these two results by pointing out that spatial correlation of gender specific adaptation effects and evoked activity during face network localizer sessions was low in their study. More generally, it should be noted that the present study maximized sensitivity to detect gender-specific differences within the core and extended face network by studying regions-of-interest (with appropriate Bonferroni correction) defined on a per-participant basis according to individual functional anatomy. Our study therefore cannot

address the question of whether outside face-sensitive regions there might exist additional cortical areas that show sensitivity to facial gender.

I further evaluated whether facial gender could be predicted from patterns of activity in visual cortex. Classification performance was comparatively poor and missed significant difference from chance, however, it was trending towards significance [Figure 3, right]. This result might indicate a possible influence of low-level stimulus properties and/or top-down feedback loops affecting the BOLD signal in early visual cortices during the duration of stimulus blocks. Assuming the existence of such top-down feedback might also help explaining the slightly higher classification accuracy results in early visual cortex as compared to other control regions.

Within the face network, the lateral FG has previously been reported to play a dominant role within the face network, indicated by consistent and replicable patterns of activation within this region, irrespective of face formats, tasks, and experimental conditions (Kanwisher et al., 1997; Ishai et al., 2000; Grill-Spector et al., 2004; Kranz and Ishai, 2006). Analysis of effective connectivity recently revealed that the FG is a major node in the face network (Fairhall & Ishai, 2007). As the FG provides the major causal input into the extended system, which processes emotional and social aspects of face stimuli, and given its pivotal role in face-perception, one might assume that gender-specific information might be presented in the FG. Our results, however, suggest that information

about face gender is a distributed attribute, rather than localized to one or two regions.

Discriminating the gender of face stimuli is an automatic and effortless task. Our results suggest that information about face gender is a distributed attribute, represented in almost all regions of the face network. Given the evolutional importance of gender information and its fundamental nature in face processing, it is plausible that there is no "gender-specific region" in the human brain, but rather, gender information is a distributed attribute as our results indicate. This conclusion is further supported by considering prosopagnosic patients who, despite their profound inability to recognize faces, exhibit normal patterns of activation in the FG (Hasson et al., 2003; Rossion et al., 2003), suggesting that activation in the FG alone is insufficient for face recognition. Taken together, previous findings and our current result suggest that face perception depends on integration of information across cortical regions and, specifically, gender information is a distributed quality within this network.

Previous studies suggest that, across multiple sensory modalities, the amygdala is reactive to very simple cues of threat or danger (Vuilleumier et al., 2003; Whalen et al., 2004). Previous work also demonstrated that schematic faces including even minimal clues of threat (i.e. eyebrows in a downward V-shape) (Wright et al., 2002) are enough to activate threat detection related activity in the amygdala. Our failure to find evidence for representations of facial gender in the amygdala cannot rule out the potential presence of

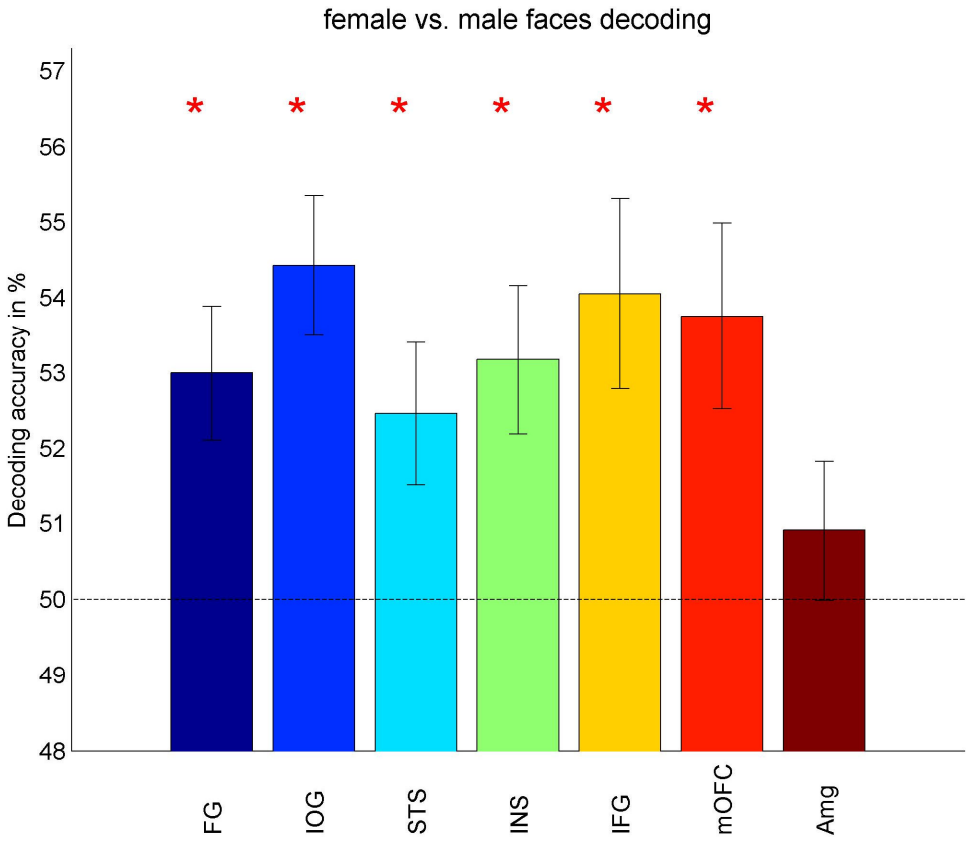
gender specific information at the level of single neurons, especially given the established role of the amygdala within the face-network (Ishai et al., 2004, Ishai et al., 2005). However, as gender discrimination is not essential for the detection of threat or danger, is may be reasonable to conclude that gender-specific information might represented in the amygdala to a lesser degree than in other regions of the face network.

ERP studies suggest that gender specific processing might occur as early as 45-85ms after the presentation of faces in face perception as part of coarse visual categorization and boosted around 200ms by attention-based gender categorization (Mouchetant-Rostaing et al., 2000). Unfortunately, Mouchetant-Rostaing and colleagues (2000) did not conduct source localization and so it is not possible to determine whether the cortical generators of the ERP effects they observed might correspond to the cortical loci I identified as encoding facial gender. In addition, the temporal limitation of fMRI data acquisition prevents us from effectively comparing the two studies. Yet, due to this temporal limitation I cannot exclude the possibility that ROI specific gender-discrimination effects as described in this study might be influenced by various top-down feedback loops affecting the BOLD signal during the duration of a stimulus block.

Despite the relative low temporal and spatial resolution, the large size of our dataset (n=40) was one of the advantages of our study. Future studies will determine whether cutting-edge high-resolution data acquisition sequences would enable higher classification

accuracies in individual subjects. Additionally, future studies might consider investigating activation patterns not only in sets of ROIs, but also in a full brain approach where the signatures of cognitive thought processes might be expressed as 'brain states' with distributed activation patterns. However, results presented here should be seen as achieved despite the limitations of the scanning procedures of the original study. It was reassuring, that, even without specifically designed task or stimuli and measuring at the low temporal resolution of fMRI, the neural representations of face perception in the face network are sufficient to conclude that gender information is represented in a distributed fashion in the human brain.

## VII.5 Supplemental Material, chapter VII



Results comparable to those reported in Figure 3 but now generated only from rating runs. Overall classification accuracies are slightly lower than in Figure 3, which might reflect the much smaller quantity of data analysed. However, the pattern of results across different areas is highly similar to those in Figure 3. This is an indirect indication that the different tasks (passive viewing vs. attractiveness rating) of the original study did not influence the main result other than by strengthening the statistical power through more training and test examples [FG: 53%, $t(39) = 3.4$, $P = 0.002$; IOG: 54.4%, $t(39) = 4.8$, $P < 0.001$; STS: 52.5%, $t(31) = 2.3$, $P = 0.03$; INS: 53.2%, $t(29) = 2.8$, $P = 0.009$; IFG: 54.1%, $t(34) = 3$, $P = 0.005$; mOFC: 53.8%, $t(25) = 2.5$, $P = 0.02$; Amg: 50.9%, $t(21) = 0.73$, n.s.].

Additionally, we also repeated classification analysis for the three passive viewing sessions. One of the difficulties with these sub-analyses is that the number of sessions per subject is substantially reduced, thus reducing power compared to the main analyses. With only 3 independent sessions, classification can only be learned on 2 sessions and tested on the third. Two sessions, however, do not offer a rich enough dataset for the SLR-classifier to generalize activity patterns specific enough to allow classification. The uncorrected results for all face-network ROIS were: FG: 51.7%, $t(38) = 2.2$, $P = 0.03$; IOG: 50.5%, $t(38) = 0.78$, P = n.s.; STS: 50.8%, $t(30) = 0.98$, P = n.s.; INS: 51.3%, $t(28) = 1.3$, P = n.s.; IFG: 50.8%, $t(33) = 0.86$, P = n.s.; mOFC: 50.7%, $t(25) = 0.82$, P = n.s.; Amg: 49.9%, $t(21) = -0.11$, P = n.s. These results are not Bonferroni corrected. Thus, classification accuracies are globally low

for passive viewing, but this is likely to be due to the significantly reduced sample size.

While we have confidence that this result is likely due to the significantly reduced sample size, we did not specifically test this assumption as the original study only contained only three passive viewing sessions. Thus, whether this reduced classification result for the three passive viewing sessions were truly non-task-related remains to be seen in future studies.

**Supplementary table 1:**

| | male vs. female participants: | participants interested in men vs. participants interested in women | homo vs. hetero participants: |
|---|---|---|---|
| IOG | (56, 54.5)<br>t(38) = 0.91, n.s. | (54.7, 53.1)<br>t(38) = -1.1, n.s. | (53.9, 56.6)<br>t(38) = 1.7, n.s. |
| FG | (54.4, 56.1)<br>t(38) = 1.3, n.s. | (53.2, 54.6)<br>t(38) = -1.1, n.s. | (55, 52.8)<br>t(38) = -1.8, n.s. |
| STS | (52.4, 54)<br>t(30) = -1.1, n.s. | (54.2, 51.8)<br>t(30) = 1.8, n.s. | (52.5, 53.9)<br>t(30) = -0.99, n.s. |
| INS | (53.4, 53)<br>t(28) = 0.21, n.s. | (52.7, 53.8)<br>t(28) = -0.55, n.s. | (53.4, 53)<br>t(28) = 0.21, n.s. |
| IFG | (54.5, 56.1)<br>t(33) = -0.88, n.s. | (54.7, 55.7)<br>t(33) = -0.55, n.s. | (55.9, 54.4)<br>t(33) = 0.86, n.s. |
| OFC | (55, 55.7)<br>t(24) = -0.34, n.s. | (56.4, 54.2)<br>t(24) = 1.1, n.s. | (56, 54.4)<br>t(24) = 0.79, n.s. |
| Amg | (51.4, 50.9)<br>t(20) = 0.32, n.s. | (51.4, 51.1)<br>t(20) = 0.14, n.s. | (51.7, 50.8)<br>t(20) = 0.5, n.s. |

Complete table of statistical values for t-tests between different subgroups of the subject population. The first two values always reflect the group means. All t-test results are Bonferroni corrected

for multiple comparisons. There were no significant differences between any of the subgroups in any of the ROIs.

**Supplementary table 2**

| | Voxels used: Mean (std) across subjects | Successful voxels Mean (std) across subjects | Voxels per sphere |
|---|---|---|---|
| IOG | 28 (4.4) | 12 (3.9) | 506 |
| FG | 32 (4.8) | 15 (4.5) | 438 |
| STS | 28 (6.4) | 15 (5.1) | 362 |
| INS | 30 (5) | 16 (3.9) | 438 |
| IFG | 29 (3.9) | 13 (3.8) | 438 |
| OFC | 28 (4.4) | 13 (3.6) | 438 |
| Amg | 25 (2.6) | 13 (2.9) | 362 |
| CTR1 | 35 (5.2) | 16 (4.6) | 1012 |
| CTR2 | 28 (3.8) | 14 (4.6) | 571 |
| CTR3 | 32 (4.7) | 16 (4.4) | 438 |
| OP | 29 (3.7) | 14 (4.7) | 362 |

Complete table of number of voxels used, successful voxels and voxels per sphere. Successful voxels were defined as voxels that where used in minimally 3 cross-validations with above chance outcome.

# Chapter VIII:

# General Discussion

## VIII.1 Introduction

The experimental studies outlined in this thesis demonstrate that feature selective processing in the brain can be studied with multivariate analysis techniques for fMRI even if features are represented at a lower spatial scale than the resolution of fMRI. Conceptually the experimental work in this thesis can be split into three distinct parts: Part one consists of chapter 3, 4 and 5 investigating the influence of varying load on feature-specific distractor processing. Part two consists of chapter 6 and investigates the influence of feature-based attention. I measured the representation of attended visual features in retinotopic areas where they were present and attended or not present (but still attended). Finally, part three is presented in chapter 7 and deals with the representation of facial gender across a network of face responsive areas. In this chapter, I will recapitulate the main findings of these three parts, as well as discuss common shortcomings, strengths and conclusions for each part and, finally, debate the scientific significance and impact of these studies. To finish, I provide an outlook on possible future studies.

### VIII.1.1 Feature specific distractor processing under load

In chapter 3, 4 and 5, I utilized a central task to manipulate attentional load (see figure III-1) while stimulating the periphery with either distractor oriented lines or distractor moving dots. Using this experimental paradigm, I was able to replicate a number of previous results, namely a behavioral (load-) manipulation with the central load task (Schwartz et al., 2005), a difference in distractor related brain activity across early visual cortices and MT (for motion only) due to this load manipulation (Rees et al., 1997; Schwartz et al., 2005) and successful feature-decoding of two types of oriented lines or two types of directions of motions (Haynes and Rees, 2005a; Kamitani and Tong, 2005b, 2005a). Apart from these replications, results showed a significant main effect of decreased univariate BOLD signal associated with irrelevant (motion- and orientation-) distractors in areas V1, V2 and V3. Additionally, I demonstrated generalization of training data in one load condition to test data taken from another load condition (cross-condition decoding). This led to the finding that pooling training data from multiple conditions for the training part of multivariate pattern decoding (MVPD) can result in more robust and more reliable decoding accuracies.

Most importantly, in all three chapters my main hypothesis was a reduction of MVPD accuracy of the visual feature contained in the distractor under high load (compared to low load). Yet, orientation-classification under high load and low load in the central task was statistically indistinguishable; the same was true for classification

performance for direction of motion. In other words, while taking the spatial pattern of activity in early visual cortices into consideration, multivariate analysis allowed accurate classification of the neural representation of one of two orientations and motion directions, but MVPD of visual features represented by the distractor were not significantly different under high load and low load. This result was not expected and possible reasons for this lack of a difference are discussed in great detail in the respective chapters (see especially chapter 3 and 4).

### VIII.1.2 Specific shortcomings and strengths of part one

Chapter 3, 4 and 5 all suffer from the same problem: the main experimental hypothesis was not confirmed and there is a distinctive lack of a good, new, interpretation of this result. In other words, even in hindsight a general alternative hypothesis does not emerge easily. This might be mainly due to the novelty and the associated lack of experience with the novel and innovative use of MVPD as a dependent measure. In recent years, MVPD has become an accepted and frequently used tool in fMRI-data analysis; however, to my knowledge, so far there are no other studies that successfully utilize *differences* in MVPD performance as a *dependent* measure to distinguish between two or more conditions. Instead, MVPD results are more used in a yes/no approach where a significant result is determined by significantly difference from chance (i.e. recent results of this nature include Formisano et al., 2008; Haushofer et al., 2008; Kay et al., 2008; Mitchell et al.,

2008; Sumner et al., 2008; Yamashita et al., 2008; Harrison and Tong, 2009; Serences et al., 2009a). Only one study utilizes MVPD to create voxel-based tuning curves (Serences et al., 2009b) and compares the effect of different attentional conditions on these tuning curves. In short, interleaved with a main experiment, the authors scan separate fMRI sessions in which a single orientation is presented on each trial, systematically varied over the full 360º of orientations across trials. Using multivariate classification, each voxel in early visual areas results in different classification performance for each orientation. Serences et al (2009) demonstrated that performance varied systematically across voxels according to different proportions of underlying orientation selective neural tissue. However, there are vast differences between paradigm and analysis in chapter 3, 4 and 5 and Serences et al (2009) which makes it hard to take results from one study to inform the other. Consequently, there is a lack of prior results, along with their interpretations, in order to precisely evaluate the result (or the lack thereof) presented here.

As an illustration, let me assume the main hypotheses in chapter 3, 4 and 5 would have been fulfilled. It could be argued that significant differences between high and low load conditions would not necessarily have been a great deal easier to interpret. Assuming significant differences in MVPD performance for irrelevant visual features as a function of central load, what conclusions would this theoretical result yield? The straightforward explanation of the result would be that increased attentional load had decreased the 'perceptual quality' of a visual feature. Yet this conclusion seems

somewhat premature as perception of the distractor was not actually measured (it was irrelevant and ignored). Thus, instead of a straightforward interpretation, further research would be necessary. However, what would this theoretical result on its own mean? Due to the lack of experience with MVPD results of ignored stimuli this question would potentially still remain elusive. However, a significant difference would have certainly hinted towards the possibility that the underlying neural activity of distractors was modulated not only in quantity (univariate result) but also in (feature-specific) quality (multivariate result). Yet, returning to reality and the actually observed null-result of no difference between decoding accuracies between different load conditions, it is important to note that this null-result does not exclude the possibility of a qualitatively altered neuronal representation due to the different load conditions. Additionally, a null-result of no significant difference does also not exclude a possible difference between the underlying neural representations of the visual features measured in the two load conditions.

However, assuming the results from chapter 3, 4 and 5 are correct and meaningful, how can be interpreted? In chapter 2, I reviewed the literature and the accepted hypothesis why multivariate classification of BOLD data works at all (for detailed discussion see chapter 2 and 4; Albright et al., 1984; Malonek et al., 1994; Tootell et al., 1995; Shmuel and Grinvald, 1996; Weliky et al., 1996; Tolias et al., 2001; Nishida et al., 2003; Seiffert et al., 2003; Haynes and Rees, 2005a; Kamitani and Tong, 2005b). I concluded that overall BOLD signal changes did not directly correspond to multivariate

results, thus, the selected voxels did not reflect trends in the overall BOLD-signal. Load-decoding results from chapter 3, 4 and 5 further corroborate this conclusion. Moreover, I speculated that selected voxels represent brain activity that was (strongly) biased towards the basic visual feature that was tested for. If the feature was represented in a biased way in some voxels, these were found by ARD and, hence, classification was successful. In the context of the results of chapter 3, 4 and 5, some voxels in V1-V3 and MT unequally represent either different directions of motion or orientations. Thus, it seems relatively certain that the existence of biased voxels is due to the existence and anisotropic distribution of feature selective columns in early visual areas and area V5/MT.

For the interpretation of the results in chapters 3, 4 and 5 this creates an interesting scenario: It is possible that load might affect the overall level of activity (i.e. the neuronal firing rates) without actually altering the tuning of those neurons. This would explain why MVPD accuracy was maintained, even under overall reduced activity levels (firing rates). Hence, those voxels biased by the direction of motion in or the orientation of the distractor remained equally biased irrelevant of the central load condition. This result must be considered carefully when trying to gain insights into the neuronal organization of human early visual cortex. By no means has it presented conclusive evidence for any specific neuronal organization; however it still represented a furthering of our understanding of the nature of attentional effects in the early visual cortex. As a consequence, the results of chapter 3, 4 and 5, under

this new interpretation, add to the understanding of the effects of attentional load on visual representations in general.

In conclusion, many behavioral measures as well as overall measures of the BOLD signal decrease under high load (Lavie and Fox, 2000; O'Connor et al., 2002; Pessoa et al., 2002; Jenkins et al., 2003; Lavie and de Fockert, 2003; Lavie, 2005; Schwartz et al., 2005; Lavie, 2006; Forster and Lavie, 2007, 2008). However, the underlying, detailed, visual feature specific representation of an irrelevant distractor might remain relatively unchanged by varying attentional load towards a central task. This speculation would mean that the attentional demands of a central task might be somewhat irrelevant for this feature specific, neuronal representation. Considering the similar results for the basic visual features of orientation and direction of motion, these results might be replicable with other basic visual features such as color or maybe even luminance – of course so far limited to feature specific neural information as read out with MVPD. However, the underlying neural causes of this conclusion are relatively open and remain largely speculative.

### VIII.1.3 Scientific novelty and significance

Experiments in chapters 3, 4 and 5 yielded a number of novel scientific conclusions and significant findings. Firstly, they represent a structured series of studies with only small changes in experimental paradigm. Thus, any conclusions derived on the basis

of all three studies are potentially highly robust and reliable: It thus seems likely that neural representations of visual features of any irrelevant peripheral distractor, when measured by MVPD, are not influenced by the amount of attentional load demanded in a central task. Chapter 3, 4 and 5 demonstrate this finding for the visual features of orientation and direction of motion. Consequently, neural representations of other visual features, such as different colors or spatial frequencies, might also remain unchanged by varying attentional load of a central task. However, these conclusions remain speculative and await formal testing.

Secondly, from results in all three chapters, I conclude that pooling training data across conditions was highly beneficial for MVPD reliability and robustness. More general, I conclude that more training data lead to a more accurate voxel-selection and, in turn, to a more accurate decoding performance. Thus, future studies utilizing MVPD to distinguish between two or more conditions containing the same or similar visual features should consider pooling training data of both conditions, training on this enlarged data set, but-testing the different conditions separately.

Thirdly, previous studies have only ever reported similar multivariate findings for *attended* motion directions and orientations (Haynes and Rees, 2005a, 2005b; Kamitani and Tong, 2005a, 2005b, 2006; Serences and Boynton, 2007b, 2007a). Results from chapter 3, 4 and 5 extend these findings to *ignored,* unattended visual features. This extension is valid under high and low load in the attended task. The finding that MVPD also provides robust

decoding for unattended stimuli, even under high load is of scientific significance in so far as it could provide the basis to further MVPD-studies of unattended stimuli.

Fourth, chapter 4 and 5 demonstrate a previously unreported significant main effect of decreased BOLD signal associated with an irrelevant (motion-) distractor in areas V1, V2 and V3. This is while also replicating previous results of an interaction of decreased motion (vs. no motion) under high (vs. low) load in V5/MT. Thus, this finding fills a gap in the knowledge and is therefore of scientific significance. Taken together with results from orientation (chapter 3) and checkerboards (Schwartz et al., 2005), it seems likely that increased attentional load in a central task decreases distractor-related BOLD activity in general. However, again, this conclusion remains speculative until it is formally tested.

### VIII.1.4 Conclusions and future directions from part one

Experimental studies in chapter 3, 4 and 5 of this thesis do not have a clear cut main conclusion. Various results within the chapters provide evidence towards conclusions and novel findings (see above) and the replications of previous studies strengthen the validity of these previous findings. Yet, the null-result for the main hypothesis in all three chapters was a disappointment due to the non-informative character of null-results and the lack of an alternative hypothesis. However, it seems likely that, the underlying visual feature specific representation of an irrelevant distractor

might remain relatively unchanged by varying attentional load towards a central task, despite overall changing activity levels. This specific interpretation awaits further testing.

Very recently, the normalization model of attention (Reynolds and Heeger, 2009) received much international acclaim. The theory proposed an elegant solution to the well known contradiction between the interaction between attention and visual stimulus contrast (see below for details). In search of new interpretations of the results from part one of this thesis, new theories like this might promise new insights. However, before jumping to conclusions it is important to notice that the normalization model of attention makes specific predictions for the neural representation of *attended* stimuli. This stands in contrast to the experimental results of chapter 3, 4 and 5 in which the focus of a participant's attention was towards a central task but the results of all three studies were derived from *unattended,* irrelevant*,* distractors. As such, the normalization theory of attention does so far not make clear predictions towards the neural representation of irrelevant distractors. A future direction of this research might thus be to extend and unify existing theories like the normalization theory of attention (Reynolds and Heeger, 2009) and the load theory of attention (Lavie et al., 2004) to include aspects of distractor related activity on a feature-specific level on neural representation.

## VIII.2 The influence of feature-based attention on unstimulated areas of the visual field

Part two of this thesis comprises chapter 6. It presented a study in which participants were presented with an initial 8 second decision period, followed by a 30s experimental block. At the beginning of the initial decision period, participants were instructed to decide, at random, between attending either a 45º or 135º oriented lines during the main task and hold their decision in mind. Then, during the main task participants attended the chosen set of oriented lines. The oriented lines appeared in a circular region in one visual quadrant only while all other visual quadrants displayed nothing but the medium gray background. Using multivariate decoding, I was able to demonstrate that BOLD activity patterns within the representation of the stimulated patch contained spatial activity patterns that could distinguish the two attentional conditions significantly better than chance. This replicated a previous finding that feature-based attention had a modulatory effect on underlying neural activity in early visual cortices (Kamitani and Tong, 2005b), however now, for stimuli contained entirely within one quadrant of the visual field.

Further, I found significantly different spatial activity patterns for the two feature-based attentional conditions across early visual areas V1, V2 and V3 even after excluding any areas of cortex that could potentially be spatially related to the stimulated area. I thus show that, in humans, attending to a particular specific feature

(here orientation) specifically modulates feature-selective units across the entire visual field independent from spatial attention. This modulation can be measured with fMRI and was found in all of early visual cortex areas V1, V2 and V3.

Investigating an 8 second period preceding the main task, during which participants had to hold a decision in mind, I demonstrated that BOLD activity from early visual cortices during that delay period were sufficient to decode which orientation a participant subsequently attended. Thus, these activity patterns contained a representation of the behavioral decision that a participant had made and was about to execute.

### VIII.2.1 Specific shortcomings and strengths of part two

One previous study has investigated the influence of feature-based attention using multivariate pattern recognition techniques, but has some important methodological shortcomings. Serences & Boynton (2007a) showed that allocating attention to one or the other of the two superimposed motion-surfaces to a stimulated patch represented in retinotopic visual areas of the contralateral hemisphere produced a differential feature-specific pattern of activation in a corresponding patch in ipsilateral retinotopic cortex representing an unattended patch, even when the second patch had received no stimulation (Serences and Boynton, 2007a). The authors claim to have demonstrated the influence of feature-based attention. However, as discussed in detail in chapter 6, this study

has numerous major shortcomings like uncorrected-for multiple statistical tests, reporting areas without observing any significant result (notably V1), incoherence of results with stimuli present and absent and potential mis-definition of area V5/MT. Therefore it remained elusive whether feature based attentional modulations could be measured with fMRI in early visual cortices.

In chapter 6, I demonstrated that attending to one of two overlapping orientations elicits differential BOLD activation patterns at the spatially attended location, at size-matched locations in other parts of the visual field and, in fact, at all other locations of the visual field as well. This showed that attending to a particular orientation in one location of the visual field specifically modulates orientation-selective units across the visual field – a much more general result than that previously demonstrated. The most likely explanation seems that the underlying neuronal populations in early visual cortices representing a particular visual feature, irrespective of retinotopic location, are modulated by feature-based attention. Moreover, such modulation can be identified even in retinotopic areas representing parts of the visual field that never receive stimulation. This result nicely parallels findings in the realm of spatial attention where anticipatory response modulations have been observed before stimulus onset (i.e. in the absence) of visual stimuli within the receptive field (Luck et al., 1997; Womelsdorf et al., 2006). However, importantly, during stimulation, spatial attention seems to be limited to the stimulated region (but potentially spreads across different features), however results from chapter 6 seems to indicate that feature-based modulation remain

intact across spatial locations (but are most likely focused on the attended feature). In the context of chapter 6, this result strengthens the hypothesis that a truly spatial-stimulus-independent feature-specific modulation was measured by MVPD fMRI.

A second potential limitation is very similar to the first one: Recently it was argued that the representation of perceptual choice between two opposite motion directions is contained in the pattern of BOLD activity of hMT+ (Serences and Boynton, 2007b). Thus, as was the case for the first set of results (above), a potential shortcoming of (decision-) results in chapter 6 could be that they could be perceived as not novel, but to replicate a previous finding with a different set of stimuli. However, as discussed in detail in chapter 6, the result by Serences and Boynton (2007b) was subsequently discussed as questionable (Kaul and Bahrami, 2008) as the experimental paradigm utilized did not control for the possibility that participants simply attended more to one direction. Should this be the case, the result by Serences and Boynton (2007a) would be a replication as it would have measured feature based attentional modulation in the very same hMT+ area that the authors showed similar modulations in the earlier study along with all of it's own shortcomings as discussed above and in chapter 6 (Serences and Boynton, 2007a). Kaul and Bahrami (2008) further suggest that this alternate explanation would also eradicate the perceived differences between the results by Serences and Boynton (2007b) and monkey studies using similar experimental procedures (Shadlen and Newsome, 2001; Williams et al., 2003).

### VIII.2.2 Scientific novelty and significance

For the first time, results in chapter 6 show that attending to a particular specific orientation (feature) specifically modulates orientation-selective (feature-selective) units across the entire visual field as measured in multiple regions of early visual cortex: V1, V2 and V3. The results demonstrate the independence of feature-based attention from spatial attention parallel to their independence proposed in monkeys (Treue and Martinez Trujillo, 1999; Martinez-Trujillo and Treue, 2004).

Additionally, results in chapter 6 demonstrate for the first time that a behavioral decision is represented in early visual areas before participants receive any visual information. As a consequence, I provide evidence that feature-based attention influences visual cortices in association with the formation of a behavioral choice, even before any attentional selection has occurred. Thus, as humans form behavioral decisions to attend to this or that feature (here orientation, but similar effects are likely for color, direction of motion, shape, facial gender and many more) their earliest visual cortices may automatically adapt to these decisions independently of whether the attended feature is presently perceived or in the focus of visual attention.

Results of this part of my thesis are consistent with key assumptions in a recently published theory of attention. The

normalization theory of attention (Reynolds and Heeger, 2009) elegantly solved the well known contradiction between the interaction between attention and visual stimulus contrast. In short, combining manipulations of contrast and attention, studies have shown that attention causes changes in either "contrast gain" or "response gain" (for review see Reynolds and Chelazzi, 2004; Carrasco, 2006). Response gain means that the neural responses are increased multiplicatively by applying a fixed "response gain" factor - thus, at any stimulus contrast attention multiplies the neural response to make it larger. Conversely, contrast gain means that attention allows lower contrast stimuli to be processed 'as if' they would be of higher contrast - thus, attention shifts the neural responses, rather than multiplying them, so that responses are larger for some but not all contrasts. Evidence consistent with both models has been empirically demonstrated by direct recording of neural responses in animals, and behavioral measurements in humans (e.g. Ling et al., 2009). Consequently, strong support exists for both of the models despite their contradictory nature, resulting in an impasse in the field of visual attention research.

The normalization model of attention elegantly solves the above contradiction by making the influence of attention dependent on the size of the stimulus and the spread of the 'attention field'. The attention field is a theoretical concept that represents how attentional feedback signals affect stimulus-evoked responses in visual cortex. According to the size of the attention field relative to the size of the stimulus, a switch from contrast to response gain is predicted. A review of previous research showed that studies

demonstrating contrast gain may have encouraged subjects (humans and animals) to utilize a large attention field, while studies demonstrating response gain may have encouraged a small attention field.

Importantly, the normalization model of attention builds upon the assumption of independence between feature-based and spatial attention. Specifically, the theoretical concept of the attention field relies on the fact that visual features and spatial location are orthogonal factors. Notably, for feature-based attention as explored in chapter 6, the theory predicts that attending to a certain feature will modulate brain activity regardless of spatial location or retinotopic representation. Results in chapter 6 demonstrate this key assumption of the normalization model of attention. As a consequence, the normalization theory of attention and results presented in chapter 6, both gain scientific significance by being mutually supportive.

### VIII.2.3 Conclusion from and future directions for part two

Results of chapter 7 are in good agreement with results from comparable studies from single cell monkey electrophysiology. This demonstrates how both approaches can inform each other and how they can and should be combined to answer new questions. Findings from this study help bridge the gap between previous electrophysiological recordings in monkeys and studies of human perception. By using functional imaging and multivariate decoding, I

demonstrate that feature-based attention modulates spatial patterns of activity in the early human visual cortex. The results further demonstrate the independence of feature-based attention from spatial attention parallel to their independence proposed in monkeys and humans (Treue and Martinez Trujillo, 1999; Martinez-Trujillo and Treue, 2004; Reynolds and Heeger, 2009). Thus, our results illustrate that feature-based attention acts to selectively enhance specific features in our visual environment resembling the currently attended set of features. This might potentially occur at the expense of information about less relevant aspects or features. However, this study also opens up new questions, for example how integration of the various types of attention identified so far (e.g. spatial and feature-based but also object or surface-based attention, etc.) might occur at the level of single neurons as well as at the level of BOLD activation patterns.

As a second result, this study presents evidence that a behavioral decision can be measured reliably in an early visual area. It thus shows an involvement of early visual cortices in behavioral decisions without these visual features necessarily visually present at the time a behavioral decision is formed. While this study noticeably demonstrates this involvement, further research is necessary to determine the exact role of early visual cortex modulation due to a behavioral decision. For example early visual areas might play an active role in the determination of a behavioral choice, or, early visual areas receive feedback input, modulating their activity as soon as a decision has been formed in a higher area, or, alas, a

combination between the two scenarios. Further research should investigate the exact role of this involvement.

### VIII.3 Facial gender representation in the human brain

Part three of this thesis comprises chapter 7. The study presented in chapter 7 sought to uncover differential brain responses associated with viewing the gender of face stimuli. Using data that was previously collected for a study looking at facial attractiveness ratings (Kranz and Ishai, 2006), I mapped face responsive brain areas of the core (fusiform gyrus (FG), inferior occipital gyrus (IOG) and superior temporal sulcus (STS)) and extended (amygdala (AMG), insula (INS), inferior frontal gyrus (IFG) and orbitofrontal cortex (OFC)) face network (Ishai et al., 2005). I demonstrated that BOLD signals averaged over these ROIs in response to seeing either female or male faces was comparable in all of these areas and was not sufficient to distinguish facial gender. However, using multivariate pattern analysis, I showed that BOLD signals from the FG, IOG, STS, INS, IFG and medial OFC all contained sufficient information to decode gender of observed faces significantly better than chance. In the amygdala and four control regions, however, significant gender classification performance was not detected. I explored the data for any differences in gender classification between different subgroups (home or hetero women and men) within the participant population. However, I did not find any significant differences in gender decoding accuracy as a function of the participant's gender (men vs. women) or their sexual

orientation (hetero- vs. homosexual) or between participants interested in men versus participants interested in women.

### VIII.3.1 Shortcomings, strengths and scientific significance of part three

One potential shortcoming could be a seeming mismatch between results in chapter 7 and a previous study reporting fMRI adaptation effects for facial gender and race outside the face network, specifically in the cingulate gyrus (Ng et al., 2006). To address this issue, I probed an anatomically defined cingulate gyrus ROI, but did not find any significant gender specific activation with conventional univariate analysis or any decoding results significantly different from chance with multivariate pattern analysis. However, mediating between the two results, Ng et al (2006) pointed out that spatial correlation of gender specific adaptation effects and evoked activity during face network localizer sessions was very low in their study.

A further potential problem lies within the face network literature itself: The lateral FG has previously been reported to play a very dominant role within the face network, indicated by consistent and replicable patterns of activation within this region, irrespective of face formats, tasks, and experimental conditions (Kanwisher et al., 1997; Ishai et al., 2000; Grill-Spector et al., 2004; Kranz and Ishai, 2006). An alternate hypothesis concerning gender processing might therefore have localized gender processing to the FG, thus strenghening the significance of the FG, alongside the IOG, within

the face network. However, our results demonstrate that information about face gender is a distributed attribute, rather than localized to one or two regions. Unfortunately, the temporal resolution of fMRI cannot address whether this distribution of facial gender information occurs through simultaneous activation of the network or a more cascade-like pattern of activation which could describe multiple levels of face processing. However, a study by Fairhall and colleagues (2007) might suggest one possibility: in that study the authors argued that the FG provides the major causal input into the extended system, which processes emotional and social aspects of face stimuli (Fairhall and Ishai, 2007). Accordingly, in that study emotional faces increased the effective connectivity between the IOG, FG, and the amygdala, whereas famous faces increased the effective connectivity between the IOG, FG, and the OFC. Thus, dynamic alterations in multiple regions of the face network depend on aspects of the task and face stimuli used. Discriminating the gender of face stimuli is an automatic and effortless task. My results demonstrate that information about face gender is a distributed attribute, represented in almost all regions of the face network. Given the evolutional importance of gender information and its fundamental nature in face processing, it is perhaps plausible that there is no "gender-specific region" in the human brain, but rather, gender information is a distributed attribute as our results indicate. This conclusion is further supported by considering prosopagnosic patients who, despite their profound inability to recognize faces, exhibit normal patterns of activation in the FG (Hasson et al., 2003; Rossion et al., 2003), suggesting that activation in the FG alone is insufficient for face recognition. Taken

together, previous findings and my current result suggest that face perception depends on integration of information across cortical regions and, specifically, gender information is a distributed quality within this network.

Previously, ERP studies suggested that gender specific processing might occur as early as 45-85ms in face perception as part of coarse visual categorization and boosted around 200ms by attention-based gender categorization (Mouchetant-Rostaing et al., 2000). Findings from chapter 7 significantly further these initial insights, despite the low temporal resolution of fMRI and even without specifically designed task or stimuli.

### VIII.3.2 Conclusion and future directions of part three

The principal conclusion of part three of this thesis is that, rather than localized to a single region, gender information seems to be a distributed attribute of face perception, represented across the core and most regions of the extended face-network. Conventional univariate analysis left an incomplete picture about the representations of specific features of face processing in the human brain. Chapter 7 is a demonstration that specific aspects of face processing can be explored with MVPD. Further research might measure the brain pattern response to different facial aspects like race, mood, age, beauty and similar attributes to get a complete picture of what aspects of face perception is computed in which regions of the brain.

Another possible direction for future research could be to combine insights from chapter 7 with the use of more natural face stimuli (Fox et al., 2008) to get better insights into the underlying neural representation of working memory. For example, such a future experiment could investigate dynamics of free viewing, cued recall and free recall. One possibility could consist of a distinctive video-clip with two or more different characters. A possible example could be a sequence between two characters in two locations of a standard TV-program. Participants would be required to watch the sequence multiple times while being scanned with fMRI. Then, in a further session, subjects could be cued towards attentively remembering certain aspects of the movie (cued recall). This could be achieved with specific questions that force subjects to think hard about a certain character. To avoid visual confounds, cueing could be visual or auditory. Participants could then be instructed to freely remember any details of the movie, indicating what they are remembering during yet another scan (free recall). In a final session subjects might see a completely new sequence involving the same characters and places.

Using MVPA and on the basis of findings from chapter 7, I hypothesise that it would be possible to decode what participants saw during free viewing to a highly specific level (such as which character but at least male/female). Then, on the basis of these brain activation patterns, I hypothesize it could be possible to discriminate which aspect of the movie (e.g. which character) has been questioned for during the cued recall session. I further

hypothesize that it would be possible to predict what subjects freely recall on the basis of brain activity alone. Finally, I speculate that it will be possible to predict what subjects saw during the final session on the basis of brain activity of previous sessions.

Potential results from such a proposed study would contribute to the understanding of face-specific visual perception, our understanding about the relationship of viewing vs. remembering (where in the brain are activation patterns preserved, where not?) and finally our understanding of freely recalling information. For the last point, especially, my hypothesis is that specific brain activation patterns will become eminent before subjects consciously report remembering a certain person or place.

## VIII.4 Final conclusion and closing remarks

In this thesis, I explored basic questions about visual processing, with one central question in mind: What are the current limits of our knowledge of brain activity underlying vision and can I further this knowledge? I have focused my experimental work around a combination of currently unknown questions about vision that can be explored, at least in part, with multivariate pattern decoding. The reason for this partly methodological focus was the heavy time-investment in the development of the MVPD-toolbox (see chapter 2). The experimental results in this thesis demonstrate that feature selective processing in the brain can be studied with MVPD. This is

true despite or especially when features are represented at a lower spatial scale than the spatial resolution of fMRI. As such, this thesis is therefore a statement suggesting the independence of MVPD from conventional fMRI analysis. Both techniques are complementary in addressing the same question: What can I infer about the underlying neural representation from an fMRI-signal? Work in this thesis has shown that in the visual system and beyond fMRI can be used as a tool to discover an anisotropic distribution of feature-based processing. As a consequence, MVPD is a powerful tool in identifying brain structures that compute a certain (visual-) feature.

Having discussed the empirical works described in this thesis, I think the most useful question for me to ask at this point is: in what way does the study of feature-specific neural processes makes me think differently in my understanding of the human brain in general or the visual system in particular? In my opinion, it seems likely that there will always remain gaps between the understanding of single cellular processing to workings of a small neuronal populations to the 'activity' difference recoded in relatively large spatial areas of the brain, like voxels. The reason is the seemingly limitless complexity of the human brain, yet it is this seemingly limitless complexity which makes the human brain so fascinating! Results in this thesis are directly furthering our understanding of the neuronal processes underlying BOLD activity in voxels as measured by fMRI. They are therefore part of many current studies unraveling the workings of the brain every day a bit further. The study of neuroscience as a whole will probably never come to a final truth

about the human brain – however this is probably just another aspect of being human.

# Chapter IX:

## References


**Aguirre GK, Zarahn E, D'Esposito M (1998) An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. Neuron 21:373-383.**

**Aharon I, Etcoff N, Ariely D, Chabris CF, O'Connor E, Breiter HC (2001) Beautiful faces have variable reward value: fMRI and behavioral evidence. Neuron 32:537-551.**

**Albright TD, Desimone R, Gross CG (1984) Columnar organization of directionally selective cells in visual area MT of the macaque. J Neurophysiol 51:16-31.**

**Allman JM, Kaas JH. (1971) A representation of the visual field in the caudal third of the middle tempral gyrus of the owl monkey (Aotus trivirgatus). Brain Res. 31:85-105**

**Allport A (1993) Attention and control: Have we been asking the wrong questions? A critical review of twenty-five years. Cambridge, MA: MIT Press.**

**Ashburner J, Friston K (1997) Multimodal image coregistration and partitioning--a unified framework. Neuroimage 6:209-217.**

**Bahrami B, Kaul C, Rees G (2009) Spatial frequency selectivity of signals in the human lateral geniculate nucleus. Neuroimage.**

**Baker CI, Hutchison TL, Kanwisher N (2007) Does the fusiform face area contain subregions highly selective for nonfaces? Nat Neurosci 10:3-4.**

**Baker JF, Petersen SE, Newsome WT, Allman JM. (1981) Visual response properties of neurons in fourextrastriate visual areas of the owl monkey (Aotus**

**trivirgatus): a quantitative comparison of medial, dorsomedial, dorsolateral, and middle temporal areas. J Neurophysiol. 45:397-416.**

**Bartels A, Logothetis NK, Moutoussis K (2008) fMRI and its interpretations: an illustration on directional selectivity in area V5/MT. Trends Neurosci 31:444-453.**

**Bichot NP, Rossi AF, Desimone R (2005) Parallel and serial neural mechanisms for visual search in macaque area V4. Science 308:529-534.**

**Breiter HC, Etcoff NL, Whalen PJ, Kennedy WA, Rauch SL, Buckner RL, Strauss MM, Hyman SE, Rosen BR (1996) Response and habituation of the human amygdala during visual processing of facial expression. Neuron 17:875-887.**

**Broadbent DE (1958) Perception and Communication. London: Pergamon.**

**Calder AJ, Young AW (2005) Understanding the recognition of facial identity and facial expression. Nat Rev Neurosci 6:641-651.**

**Calder AJ, Nummenmaa L (2007) Face cells: separate processing of expression and gaze in the amygdala. Curr Biol 17:R371-372.**

**Carl C (2004) Kernels for Structures. Publications of the Institute of Cognitive Science, University of Osnabrück 9.**

**Carrasco M (2006) Covert attention increases contrast sensitivity: Psychophysical, neurophysiological and neuroimaging studies. Prog Brain Res 154:33-70.**

**Cartwright-Finch U, Lavie N (2007) The role of perceptual load in inattentional blindness. Cognition 102:321-340.**

**Chiu Y-C, Esterman M, Rosen H, Yantis S (2009) Decoding task-based attentional modulation in the cortical face**

**network. In: Vision Sciences Society.**

**Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19:261-270.**

**Cristianini N, Shawe-Taylor, J. (2000) An Introduction to Support Vector Machines and Other Kernel-based Learning Methods: Cambridge University Press.**

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

**Deutsch JA, Deutsch D (1963) Some theoretical considerations. Psychol Rev 70:80-90.**

**Diamond R, Carey S (1986) Why faces are and are not special: an effect of expertise. J Exp Psychol Gen 115:107-117.**

**Dougherty RF, Koch VM, Brewer AA, Fischer B, Modersitzki J, Wandell BA (2003) Visual field representations and locations of visual areas V1/2/3 in human visual cortex. J Vis 3:586-598.**

**Driver J (2001) A selective review of selective attention research from the past century. Br J Psychol 92:53-78.**

**Dukelow SP, DeSouza JF, Culham JC, van den Berg AV, Menon RS, Vilis T (2001) Distinguishing subregions of the human MT+ complex using visual fields and pursuit eye movements. J Neurophysiol 86:1991-2000.**

**Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994) fMRI of human visual cortex. Nature 369:525.**

**Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. Nature 392:598-601.**

**Fairhall SL, Ishai A (2007) Effective connectivity within the distributed cortical network for face perception. Cereb Cortex 17:2400-2406.**

**Farah MJ, Wilson KD, Drain M, Tanaka JN (1998) What is "special" about face perception? Psychol Rev 105:482-498.**

**Formisano E, De Martino F, Bonte M, Goebel R (2008) "Who" is saying "what"? Brain-based decoding of human voice and speech. Science 322:970-973.**

**Forster S, Lavie N (2007) High perceptual load makes everybody equal: eliminating individual differences in distractibility with load. Psychol Sci 18:377-381.**

**Forster S, Lavie N (2008) Failures to ignore entirely irrelevant distractors: the role of load. J Exp Psychol Appl 14:73-83.**

**Fox CJ, Iaria G, Barton JJ (2008) Defining the face processing network: Optimization of the functional localizer in fMRI. Hum Brain Mapp.**

**Fox PT, Raichle ME (1986) Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. Proc Natl Acad Sci U S A 83:1140-1144.**

**Friston KJ, Josephs O, Rees G, Turner R (1998a) Nonlinear event-related responses in fMRI. Magn Reson Med 39:41-52.**

**Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R (1996) Movement-related effects in fMRI time-series. Magn Reson Med 35:346-355.**

**Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998b) Event-related fMRI: characterizing differential responses. Neuroimage 7:30-40.**

**Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R (1995) Analysis of fMRI time-series revisited. Neuroimage 2:45-53.**

**Gauthier I, Tarr MJ (2002) Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. J Exp Psychol Hum Percept Perform 28:431-446.**

**Gobbini MI, Haxby JV (2007) Neural systems for recognition of familiar faces. Neuropsychologia 45:32-41.**

**Grill-Spector K, Malach R (2004) The human visual cortex. Annu Rev Neurosci 27:649-677.**

**Grill-Spector K, Knouf N, Kanwisher N (2004) The fusiform face area subserves face perception, not generic within-category identification. Nat Neurosci 7:555-562.**

**Grill-Spector K, Sayres R, Ress D (2006) High-resolution imaging reveals highly selective nonface clusters in the fusiform face area. Nat Neurosci 9:1177-1185.**

**Grill-Spector K, Kushnir T, Hendler T, Edelman S, Itzchak Y, Malach R (1998) A sequence of object-processing stages revealed by fMRI in the human occipital lobe. Hum Brain Mapp 6:316-328.**

**Hagenaar R, van der Heijden AH (1986) Target-noise separation in visual selective attention. Acta Psychol (Amst) 62:161-176.**

**Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. Nature 458:632-635.**

**Hasson U, Avidan G, Deouell LY, Bentin S, Malach R (2003) Face-selective activation in a congenital prosopagnosic subject. J Cogn Neurosci 15:419-431.**

**Haushofer J, Livingstone MS, Kanwisher N (2008) Multivariate patterns in object-selective cortex**

**dissociate perceptual and physical shape similarity. PLoS Biol 6:e187.**

**Haxby JV, Hoffman EA, Gobbini MI (2000) The distributed human neural system for face perception. Trends Cogn Sci 4:223-233.**

**Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425-2430.**

**Haynes JD, Rees G (2005a) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nat Neurosci 8:686-691.**

**Haynes JD, Rees G (2005b) Predicting the stream of consciousness from activity in human visual cortex. Curr Biol 15:1301-1307.**

**Haynes JD, Rees G (2006) Decoding mental states from brain activity in humans. Nat Rev Neurosci 7:523-534.**

**Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007) Reading hidden intentions in the human brain. Curr Biol 17:323-328.**

**Heeger DJ, Boynton GM, Demb JB, Seidemann E, Newsome WT (1999) Motion opponency in visual cortex. J Neurosci 19:7162-7174.**

**Herrmann K, Carrasco M, Heeger DJ (2009) Psychophysical evidence for the normalization model of attention. In: Vision Sciences Society. Napels.**

**Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106-154.**

**Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. J Physiol 195:215-243.**

**Huk AC, Dougherty RF, Heeger DJ (2002) Retinotopy and functional subdivision of human areas MT and MST. J Neurosci 22:7195-7205.**

**Ishai A (2007) Sex, beauty and the orbitofrontal cortex. Int J Psychophysiol 63:181-185.**

**Ishai A (2008) Let's face it: it's a cortical network. Neuroimage 40:415-419.**

**Ishai A, Ungerleider LG, Haxby JV (2000) Distributed neural systems for the generation of visual images. Neuron 28:979-990.**

**Ishai A, Schmidt CF, Boesiger P (2005) Face perception is mediated by a distributed cortical network. Brain Res Bull 67:87-93.**

**Ishai A, Pessoa L, Bikle PC, Ungerleider LG (2004) Repetition suppression of faces is modulated by emotion. Proc Natl Acad Sci U S A 101:9827-9832.**

**Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV (1999) Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci U S A 96:9379-9384.**

**Jenkins R, Lavie N, Driver J (2003) Ignoring famous faces: category-specific dilution of distractor interference. Percept Psychophys 65:298-309.**

**Jenkins R, Lavie N, Driver J (2005) Recognition memory for distractor faces depends on attentional load at exposure. Psychon Bull Rev 12:314-320.**

**Kamitani Y, Tong F (2005a) Decoding motion direction from activity in human visual cortex. Journal of Vision 152a.**

**Kamitani Y, Tong F (2005b) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679-685.**

**Kamitani Y, Tong F (2006) Decoding seen and attended motion directions from activity in the human visual cortex. Curr Biol 16:1096-1102.**

**Kanai R, Tsuchiya N, Verstraten FA (2006) The scope and limits of top-down attention in unconscious visual processing. Curr Biol 16:2332-2336.**

**Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302-4311.**

**Kaul C, Bahrami B (2008) Subjective experience of motion or attentional selection of a categorical decision. J Neurosci 28:4110-4112.**

**Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352-355.**

**Kim DS, Duong TQ, Kim SG (2000) High-resolution mapping of iso-orientation columns by fMRI. Nat Neurosci 3:164-169.**

**Kim SG, Fukuda M (2008) Lessons from fMRI about mapping cortical columns. Neuroscientist 14:287-299.**

**Kranz F, Ishai A (2006) Face perception is modulated by sexual preference. Curr Biol 16:63-68.**

**Kriegeskorte N, Goebel R, Bandettini P (2006) Information-based functional brain mapping. Proc Natl Acad Sci U S A 103:3863-3868.**

**Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009) Circular analysis in systems neuroscience: the dangers of double dipping. Nat Neurosci 12:535-540.**

**Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R, et al. (1992) Dynamic magnetic resonance**

**imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci U S A 89:5675-5679.**

**Lavie N (1995) Perceptual load as a necessary condition for selective attention. J Exp Psychol Hum Percept Perform 21:451-468.**

**Lavie N (1997) Visual feature integration and focused attention: response competition from multiple distractor features. Percept Psychophys 59:543-556.**

**Lavie N (2005) Distracted and confused? selective attention under load. Trends Cogn Sci 9:75-82.**

**Lavie N (2006) The role of perceptual load in visual awareness. Brain Res 1080:91-100.**

**Lavie N, Tsal Y (1994) Perceptual load as a major determinant of the locus of selection in visual attention. Percept Psychophys 56:183-197.**

**Lavie N, Cox S (1997) On the efficiency of attentional selection: Efficient visual search results in inefficient rejection of distraction. Psychological Science 8:395-398.**

**Lavie N, Fox E (2000) The role of perceptual load in negative priming. J Exp Psychol Hum Percept Perform 26:1038-1052.**

**Lavie N, de Fockert JW (2003) Contrasting effects of sensory limits and capacity limits in visual selective attention. Percept Psychophys 65:202-212.**

**Lavie N, Hirst A, de Fockert JW, Viding E (2004) Load theory of selective attention and cognitive control. J Exp Psychol Gen 133:339-354.**

**Leveroni CL, Seidenberg M, Mayer AR, Mead LA, Binder JR, Rao SM (2000) Neural systems underlying the recognition of familiar and newly learned faces. J Neurosci 20:878-886.**

**Ling S, Liu T, Carrasco M (2009) How spatial and feature-based attention affect the gain and tuning of population responses. Vision Res 49:1194-1204.**

**Liu J, Harris A, Kanwisher N (2009) Perception of Face Parts and Face Configurations: An fMRI Study. J Cogn Neurosci.**

**Liu T, Larsson J, Carrasco M (2007) Feature-based attention modulates orientation-selective responses in human visual cortex. Neuron 55:313-323.**

**Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453:869-878.**

**Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. Annu Rev Physiol 66:735-769.**

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

**Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J Neurophysiol 77:24-42.**

**Mack AR, I. (1998) Inattentional Blindness: Cambridge, MA: MIT Press.**

**Malach R, Reppas JB, Benson RR, Kwong KK, Jiang H, Kennedy WA, Ledden PJ, Brady TJ, Rosen BR, Tootell RB (1995) Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. Proc Natl Acad Sci U S A 92:8135-8139.**

**Malonek D, Tootell RB, Grinvald A (1994) Optical imaging reveals the functional architecture of neurons processing shape and motion in owl monkey area MT. Proc Biol Sci 258:109-119.**

**Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. Curr Biol 14:744-751.**

**Maunsell JH, Treue S (2006) Feature-based attention in visual cortex. Trends Neurosci 29:317-322.**

**Mendoza D, Kaul C, Martinez-Trujillo JC (2009) Working memory and feature-based attention independently modulate the perception of coherent motion in human observers. In: Vision Sciences Society 2009. Naples. Florida.**

**Mitchell TM, Hutchinson R, Just MA, Niculescu RS, Pereira F, Wang X (2003) Classifying instantaneous cognitive states from FMRI data. AMIA Annu Symp Proc:465-469.**

**Mitchell TM, Shinkareva SV, Carlson A, Chang KM, Malave VL, Mason RA, Just MA (2008) Predicting human brain activity associated with the meanings of nouns. Science 320:1191-1195.**

**Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. Science 229:782-784.**

**Motter BC (1994) Neural correlates of attentive selection for color or luminance in extrastriate area V4. J Neurosci 14:2178-2189.**

**Mouchetant-Rostaing Y, Giard MH, Bentin S, Aguera PE, Pernier J (2000) Neurophysiological correlates of face gender processing in humans. Eur J Neurosci 12:303-310.**

**Mountcastle VB (1997) The columnar organization of the neocortex. Brain 120 (Pt 4):701-722.**

**Mur M, Bandettini PA, Kriegeskorte N (2009) Revealing representational content with pattern-information**

**fMRI--an introductory guide. Soc Cogn Affect Neurosci 4:101-109.**

**Neill WT (1977) Inhibitory and facilitatory processes in selective attention. Journal of Experimental Psychology: Human Perception and Performance:444-450.**

**Ng M, Ciaramitaro VM, Anstis S, Boynton GM, Fine I (2006) Selectivity for the configural cues that identify the gender, ethnicity, and identity of faces in human cortex. Proc Natl Acad Sci U S A 103:19552-19557.**

**Nishida S, Sasaki Y, Murakami I, Watanabe T, Tootell RB (2003) Neuroimaging of direction-selective mechanisms for second-order motion. J Neurophysiol 90:3242-3254.**

**O'Connor DH, Fukui MM, Pinsk MA, Kastner S (2002) Attention modulates responses in the human lateral geniculate nucleus. Nat Neurosci 5:1203-1209.**

**O'Doherty J, Winston J, Critchley H, Perrett D, Burt DM, Dolan RJ (2003) Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. Neuropsychologia 41:147-155.**

**O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially distributed representations of objects and faces in ventral temporal cortex. J Cogn Neurosci 17:580-590.**

**Ogawa S, Lee TM, Nayak AS, Glynn P (1990) Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn Reson Med 14:68-78.**

**Pauling L, Coryell CD (1936) The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. Proc Natl Acad Sci U S A 22:210-216.**

**Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. Neuroimage 45:S199-209.**

**Pessoa L, McKenna M, Gutierrez E, Ungerleider LG (2002) Neural processing of emotional faces requires attention. Proc Natl Acad Sci U S A 99:11458-11463.**

**Pinsk MA, Doniger GM, Kastner S (2004) Push-pull mechanism of selective attention in human extrastriate cortex. J Neurophysiol 92:622-629.**

**Pitcher D, Walsh V, Yovel G, Duchaine B (2007) TMS evidence for the involvement of the right occipital face area in early face processing. Curr Biol 17:1568-1573.**

**Puce A, Allison T, Gore JC, McCarthy G (1995) Face-sensitive regions in human extrastriate cortex studied by functional MRI. J Neurophysiol 74:1192-1199.**

**Puce A, Allison T, Bentin S, Gore JC, McCarthy G (1998) Temporal cortex activation in humans viewing eye and mouth movements. J Neurosci 18:2188-2199.**

**Rees G, Lavie N (2001) What can functional imaging reveal about the role of attention in visual awareness? Neuropsychologia 39:1343-1353.**

**Rees G, Frith CD, Lavie N (1997) Modulating irrelevant motion perception by varying attentional load in an unrelated task. Science 278:1616-1619.**

**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.**

**Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. Annu Rev Neurosci 27:611-647.**

**Reynolds JH, Heeger DJ (2009) The normalization model of attention. Neuron 61:168-185.**

**Rock I, Gutman D (1981) The effect of inattention on form perception. J Exp Psychol Hum Percept Perform 7:275-285.**

**Rossion B, Caldara R, Seghier M, Schuller AM, Lazeyras F, Mayer E (2003) A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. Brain 126:2381-2395.**

**Rothman SL (1998) Magnetic resonance imaging (MRI) scans. Spine (Phila Pa 1976) 23:642-643.**

**Saenz M, Buracas GT, Boynton GM (2002) Global effects of feature-based attention in human visual cortex. Nat Neurosci 5:631-632.**

**Sasaki Y, Rajimehr R, Kim BW, Ekstrom LB, Vanduffel W, Tootell RB. (2006) The radial bias: a different slant on visual orientation sensitivity in human and nonhuman primates. Neuron. 51:661-70.**

**Schwartz S, Vuilleumier P, Hutton C, Maravita A, Dolan RJ, Driver J (2005) Attentional load and sensory competition in human vision: modulation of fMRI responses by load at fixation during task-irrelevant stimulation in the peripheral visual field. Cereb Cortex 15:770-786.**

**Seiffert AE, Somers DC, Dale AM, Tootell RB (2003) Functional MRI studies of human visual motion perception: texture, luminance, attention and after-effects. Cereb Cortex 13:340-349.**

**Serences JT, Yantis S (2006) Selective visual attention and perceptual coherence. Trends Cogn Sci 10:38-45.**

**Serences JT, Boynton GM (2007a) Feature-based attentional modulations in the absence of direct visual stimulation. Neuron 55:301-312.**

**Serences JT, Boynton GM (2007b) The representation of behavioral choice for motion in human visual cortex. J Neurosci 27:12893-12899.**

**Serences JT, Ester EF, Vogel EK, Awh E (2009a) Stimulus-specific delay activity in human primary visual cortex. Psychol Sci 20:207-214.**

**Serences JT, Saproo S, Scolari M, Ho T, Muftuler LT (2009b) Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. Neuroimage 44:223-231.**

**Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889-893.**

**Seymour KJ, Scott McDonald J, Clifford CW (2009) Failure of colour and contrast polarity identification at threshold for detection of motion and global form. Vision Res 49:1592-1598.**

**Shadlen MN, Newsome WT (2001) Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 86:1916-1936.**

**Shmuel A, Grinvald A (1996) Functional organization for direction of motion and its relationship to orientation maps in cat area 18. J Neurosci 16:6945-6964.**

**Simmons WK, Bellgowan PS, Martin A (2007) Measuring selectivity in fMRI data. Nat Neurosci 10:4-5.**

**Simons DJ, Chabris CF (1999) Gorillas in our midst: Sustained inattentional blindness for dynamic events. Perception 28:1059-1074.**

**Singh KD, Smith AT, Greenlee MW (2000) Spatiotemporal frequency and direction sensitivities of human visual areas measured using fMRI. Neuroimage 12:550-564.**

**Squire LK, Novelline RA (1997) Squire's fundamentals of radiology (5th ed). Cambridge: Harvard University Press.**

**Sterzer P, Haynes JD, Rees G (2008) Fine-scale activity patterns in high-level visual areas encode the category of invisible objects. J Vis 8:10 11-12.**

**Sumner P, Anderson EJ, Sylvester R, Haynes JD, Rees G (2008) Combined orientation and colour information in human V1 for both L-M and S-cone chromatic axes. Neuroimage 39:814-824.**

**Teo PC, Sapiro G, Wandell BA (1997) Creating connected representations of cortical gray matter for functional MRI visualization. IEEE Trans Med Imaging 16:852-863.**

**Tipper SP (1985) The negative priming effect: inhibitory priming by ignored objects. Q J Exp Psychol A 37:571-590.**

**Tolias AS, Smirnakis SM, Augath MA, Trinath T, Logothetis NK (2001) Motion processing in the macaque: revisited with functional magnetic resonance imaging. J Neurosci 21:8594-8601.**

**Tootell RB, Taylor JB (1995) Anatomical evidence for MT and additional cortical visual areas in humans. Cereb Cortex 5:39-55.**

**Tootell RB, Dale AM, Sereno MI, Malach R (1996) New images from human visual cortex. Trends Neurosci 19:481-489.**

**Tootell RB, Reppas JB, Kwong KK, Malach R, Born RT, Brady TJ, Rosen BR, Belliveau JW (1995) Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J Neurosci 15:3215-3230.**

**Torralbo A, Beck DM (2008) Perceptual-load-induced selection as a result of local competitive interactions in visual cortex. Psychol Sci 19:1045-1050.**

**Tournoux JTaP (1988) Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System - an Approach to Cerebral Imaging. New York: Thieme Medical Publishers.**

**Treisman A (2004) Psychological Issues in Selective Attention. In: The cognitive Neurosciences (Gazzaniga M, ed), pp 527-545.**

**Treisman A, Geffen G (1968) Selective attention and cerebral dominance in perceiving and responding to speech messages. Q J Exp Psychol 20:139-150.**

**Treue S, Maunsell JH (1996) Attentional modulation of visual motion processing in cortical areas MT and MST. Nature 382:539-541.**

**Treue S, Martinez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. Nature 399:575-579.**

**Tsao DY, Livingstone MS (2008) Mechanisms of face perception. Annu Rev Neurosci 31:411-437.**

**Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. Science 311:670-674.**

**Tsao DY, Freiwald WA, Knutsen TA, Mandeville JB, Tootell RB (2003) Faces and objects in macaque cerebral cortex. Nat Neurosci 6:989-995.**

**Vanzetta I, Grinvald A (1999) Increased cortical oxidative metabolism due to sensory stimulation: implications for functional brain imaging. Science 286:1555-1558.**

**Vuilleumier P, Armony JL, Driver J, Dolan RJ (2003) Distinct spatial frequency sensitivities for processing faces and emotional expressions. Nat Neurosci 6:624-631.**

**Vuilleumier P, Sagiv N, Hazeltine E, Poldrack RA, Swick D, Rafal RD, Gabrieli JD (2001) Neural fate of seen and unseen faces in visuospatial neglect: a combined event-related functional MRI and event-related potential study. Proc Natl Acad Sci U S A 98:3495-3500.**

**Wandell BA, Chial S, Backus BT (2000) Visualization and measurement of the cortical surface. J Cogn Neurosci 12:739-752.**

**Wang G, Tanaka K, Tanifuji M (1996) Optical imaging of functional organization in the monkey inferotemporal cortex. Science 272:1665-1668.**

**Wang G, Tanifuji M, Tanaka K (1998) Functional architecture in monkey inferotemporal cortex revealed by in vivo optical imaging. Neurosci Res 32:33-46.**

**Weliky M, Bosking WH, Fitzpatrick D (1996) A systematic map of direction preference in primary visual cortex. Nature 379:725-728.**

**Whalen PJ, Kagan J, Cook RG, Davis FC, Kim H, Polis S, McLaren DG, Somerville LH, McLean AA, Maxwell JS, Johnstone T (2004) Human amygdala responsivity to masked fearful eye whites. Science 306:2061.**

**Williams ZM, Elfar JC, Eskandar EN, Toth LJ, Assad JA (2003) Parietal activity and the perceived direction of ambiguous apparent motion. Nat Neurosci 6:616-623.**

**Womelsdorf T, Anton-Erxleben K, Pieper F, Treue S (2006) Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. Nat Neurosci 9:1156-1160.**

**Wright CI, Martis B, Shin LM, Fischer H, Rauch SL (2002) Enhanced amygdala responses to emotional versus**

**neutral schematic facial expressions. Neuroreport 13:785-790.**

**Yamashita O, Sato MA, Yoshioka T, Tong F, Kamitani Y (2008) Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns. Neuroimage 42:1414-1429.**

**Yi DJ, Chun MM (2005) Attentional modulation of learning-related repetition attenuation effects in human parahippocampal cortex. J Neurosci 25:3593-3600.**

**Zeki S (1983) The distribution of wavelength and orientation selective cells in different areas of monkey visual cortex. Proc R Soc Lond B Biol Sci 217:449-470.**

**Zeki S, Watson JD, Lueck CJ, Friston KJ, Kennard C, Frackowiak RS (1991) A direct demonstration of functional specialization in human visual cortex. J Neurosci 11:641-649.**