Contextual Modulations of Visual Perception and Visual Cortex Activity in Humans

Benjamin de Haas

Supervisors:

Prof Geraint Rees

Prof Vincent Walsh

Dissertation submitted for the degree of

Doctor of Philosophy, University College London

Institute of Cognitive Neuroscience

Wellcome Trust Centre for Neuroimaging

Declaration

I, Benjamin de Haas, 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 the thesis.

The experiments described in Chapter 3 were carried out in collaboration with Prof Jon Driver's group at the Wellcome Trust Centre for Neuroimaging, UCL. Dr. Ryota Kanai (ICN, UCL) kindly shared neuroanatomical data for the study described in chapter 4. Dr. Nikolaus Kriegeskorte and Dr. Linda Henriksson from the MRC-CBU in Cambridge kindly shared the stimuli used in the study described in chapter 7.

The work presented in Chapters 3, 4, 5, and 6 has been published as the following papers:

Romei V, De Haas B, Mok RM and Driver J (2011). *Auditory stimulus timing influences perceived duration of co-occurring visual stimuli*. Front. Psychology **2**:215. (Chapter 3)

de Haas B, Cecere R, Cullen H, Driver J, Romei V (2013). *The Duration of a Co-Occurring Sound Modulates Visual Detection Performance in Humans.* PLoS ONE 8(1): e54789. (Chapter 3)

de Haas B, Kanai R, Jalkanen L, Rees G (2012). *Grey matter volume in early humanvisual cortex predicts proneness to the sound-induced flash illusion*. Proc Biol Sci. 279(1749):4955-61. (Chapter 4)

de Haas B, Schwarzkopf DS, Urner M, Rees G (2013). *Auditory modulation of visual stimulus encoding in human retinotopic cortex*. NeuroImage, 70: 258‑267. (Chapter 5)

de Haas B, Schwarzkopf DS, Anderson EJ, Rees G (2014). *Perceptual load affects spatial tuning of neuronal populations in human early visual cortex*. Current Biology, 24(2): R66-R67. (Chapter 6)

Thank you!

Geraint: Thank you for granting me a fantastic combination of freedom and support, for making this a fun experience and for teaching me about the importance of 'biological questions' and 'how to become happy in science'. I'm deeply impressed and hope your supervision will leave a lasting mark on me.

Ryota, **Vincenzo** and most of all **Sam**! Thank you for being generous with time and code and for taking my ideas seriously. Sam, I learnt a ton from you and I look forward to continue doing so.

Maren & **Joe**: For all the hours of chatting, discussing, laughing and sometimes crying. I miss G07!

The Rees lab: For all the crazy discussions, for burnouts at Florida stoplights and for hanging out on the Great Wall of China.

The Wellcome Trust and 4-year PhD board: For giving me this unique opportunity. With special thanks to Sarah-Jayne Blakemore, Bertram Walter, Karen Zentgraf, Axel Kohler, Gerhardt Roth and everyone else who supported me on my way to London.

Mum, **Dad** & **Grandpa**: For all your support and interest in what I do and for teaching me curiousity by example.

Finn & **Mika**: For being the fabulous human beings you are and for constantly reminding me of what's really important.

And last, but most importantly, **Mone**: For being crazy enough to do this. For your courage, patience, strength and love. For seeing the bigger picture and putting things into perspective. For giving me the freedom and support to find out who I am. For filling our home with color, music and laughter. Thank you for being you in all the ways you are. I'm a lucky guy.

Abstract

Visual perception and neural processing depend on more than retinal stimulation alone. They are modulated by contextual factors like cross-modal input, the current focus of attention or previous experience. In this thesis I investigate ways in which these factors affect vision.

A first series of experiments investigates how co-occurring sounds modulate vision, with an emphasis on temporal aspects of visual processing. In three behavioral experiments I find that participants are unable to ignore the duration of co-occurring sounds when giving visual duration judgments. Furthermore, prolonged sound duration goes along with improved detection sensitivity for visual stimuli and thus extends beyond duration judgments per se. I go on to test a cross-modal illusion in which the perceived number of flashes in a rapid series is affected by the number of co-occurring beeps (the sound-Induced flash illusion). Combining data from structural magnetic resonance imaging (MRI) and a behavioral experiment I find that individual proneness to this illusion is linked with less grey matter volume in early visual cortex. Finally, I test how co-occurring sounds affect the cortical representation of more natural visual stimuli. A functional MRI (fMRI) experiment investigates patterns of activation evoked by short video clips in visual areas V1-3. The trial-by-trial reliability of such patterns is reduced for videos accompanied by mismatching sounds.

Turning from cross-modal effects to more intrinsic sources of contextual modulation I test how attention affects visual representations in V1-3. Using fMRI and population receptive field (pRF) mapping I find that

high perceptual load at fixation renders spatial tuning for the surrounding visual field coarser and goes along with pRFs being radially repelled.

In a final behavioral and fMRI experiment I find that the perception of face features is modulated by retinal stimulus location. Eye and mouth stimuli are recognized better, and evoke more discriminable patterns of activation in face sensitive patches of cortex, when they are presented at canonical locations.

Taken together, these experiments underscore the importance of contextual modulation for vision, reveal some previously unknown such factors and point to possible neural mechanisms underlying them. Finally, they argue for an understanding of vision as a process using all available cues to arrive at optimal estimates for the causes of sensory events.

Table of Contents

List of Figures

List of Tables

1 General Introduction

1.1 Preface

The aim of this thesis is to investigate contextual modulations of visual perception and visual cortex activity. It is a well-known fact among vision scientists that visual perception depends on more than retinal stimulation alone (Wandell, 1995). This is somewhat at odds with the phenomenological quality of vision. Visual perception feels immediate, like a direct access to the surroundings. The phrase 'I saw it with my own eyes' is used to underscore certainty and to claim objective evidence. William James put it this way:

"[…] *in both sensation and perception we perceive the fact as an immediately present outboard reality, and this makes them differ from 'thought' and 'conception', whose objects do not appear present in the immediate physical way.*" (James, 1890).

There is an interesting divergence between this strong impression of objectivity and some aspects of vision discovered by its scientific study. Seeing the world 'with our own eyes' means seeing it through a very specific filter. The visual system will encode only a certain part of the electromagnetic spectrum, it will represent only certain features of a stimulus and it will interpret those features in very specific ways (Wandell, 1995). But even identical retinal stimuli are not guaranteed to yield identical percepts. Visual perception and visual cortex activity can be modulated by contextual factors. For example, perceived visual aspects of a stimulus can change due to co-occurring *auditory* input (Shams, Kamitani, & Shimojo,

2000). A sufficiently difficult task can *distract* observers from seeing what is right in front of them (Cartwright-Finch & Lavie, 2007). And the way a stimulus looks can be subject to *spatial heterogeneity*, i.e. depend on which part of the retina was stimulated (Afraz, Pashkam, & Cavanagh, 2010).

It is this proneness of visual perception to contextual modulation that is the common theme of experiments presented in this thesis. Specifically, the first three experimental chapters investigate open questions regarding auditory modulation of vision. The focus of research on audiovisual perception has somewhat shifted in recent years. A previous view of general visual dominance (Colavita, 1974; Posner, Nissen, & Klein, 1976) has been challenged by the observation that audition can radically change visual perception (Shams et al., 2000). Newer theories of cue integration propose that the weighting of sensory channels depends on their relative levels of signal-to-noise (Ernst & Banks, 2002; Shams, Ma, & Beierholm, 2005). At the same time understanding of multisensory cortical processing has changed. Multisensory convergence can no longer be viewed as restricted to dedicated 'higher' cortical areas and has instead been shown to also affect processing in early sensory areas (Driver & Noesselt, 2008). Here I try to further investigate auditory modulation of visual perception and visual cortex processing by posing the following questions:

- Can the duration of a visual percept be modulated by the duration of a co-occurring auditory stimulus? (Chapter 3)
- Are there systematic differences in brain morphology between people more or less prone to an audiovisual illusion? (Chapter 4)

- Can co-occurring sounds modulate the accuracy with which we can decode stimuli from brain activity in V1-3? (Chapter 5)

The following chapter deals with an open question regarding attentional modulation of visual cortex activity. Attention, and specifically attentional load, has been shown to modulate the amplitude of neural responses in early visual cortex to task-irrelevant stimuli (Schwartz et al., 2005). However, it is unknown whether this modulation goes beyond mere amplitude effects; e.g. whether the spatial tuning, of neural populations is affected as well. Such effects have been observed in monkey studies investigating the effects of spatial attention (Womelsdorf, Anton-Erxleben, Pieper, & Treue, 2006), but have not been addressed for humans and attentional load before. Using the relatively recent method of population receptive field modelling (Dumoulin & Wandell, 2008) allowed me to address this question directly and non-invasively in humans:

- Does perceptual load modulate the spatial tuning of neural populations in early visual cortex? (Chapter 6)

The study of visual field maps in early visual cortex led me to wonder whether spatial preferences could play a functional role in higher visual areas as well. Specifically, I speculated that a consistent spatial input bias for a given stimulus (i.e. a canonical retinotopic stimulus location) could be reflected in spatial heterogeneity of perceptual sensitivity for the respective stimulus. Furthermore, neural populations tuned for such a stimulus might

show a corresponding bias in their spatial preference. I aimed to test this idea in two exploratory studies, asking

- Is perceptual and neural sensitivity to eye and mouth stimuli subject to spatial heterogeneity? (Chapter 7)

The aim of the general introduction is to give a brief overview of the relevant literature and to highlight open questions regarding these three topics: Contextual modulation of visual processing via co-occurring auditory stimuli, perceptual load and spatial heterogeneity in face perception.

1.2 Auditory modulation of vision

1.2.1 Auditory modulation of visual perception

1.2.1.1 Audiovisual integration

Visual input tends to dominate perceptual judgments regarding the location of an audiovisual stimulus (Pick, Warren, & Hay, 1969). Traditionally, this 'ventriloquist effect' has been interpreted as an example of a general tendency for 'visual capture' or 'visual dominance' over other sensory input (Posner et al., 1976). More recently, multisensory cue combination has been shown to be more sophisticated. The weighting of input channels appears to be reliability-weighted and thus statistically optimal (Ernst & Banks, 2002; Welch, DutionHurt, & Warren, 1986). So for example the ventriloquist effect can be reversed and spatial judgments be dominated by auditory input if the visual input is blurred (Alais & Burr, 2004). In light of these findings the usually observed visual dominance for

spatial judgments appears to simply reflect the fact that the visual system typically has higher spatial resolution than the auditory system (Witten & Knudsen, 2005).

In line with this perspective, audition tends to dominate vision in the temporal domain for which it is usually more reliable (Grondin, 2010). For example, the perceived temporal frequency of visual flicker can be 'driven' by the frequency of co-occurring auditory flutter (Gebhard & Mowbray, 1959; Shipley, 1964; Welch et al., 1986). Furthermore the perceived onset of a flash can be biased towards the onset of a co-occurring sound (Fendrich & Corballis, 2001; Morein-Zamir, Soto-Faraco, & Kingstone, 2003). Another line of research investigated auditory influences on the perceived duration of co-occurring visual stimuli (Donovan, Lindsay, & Kingstone, 2004; Klink, Montijn, & Van Wezel, 2011). These studies will be discussed in more detail below, as they are directly relevant for some of the experiments presented in this thesis. But beforehand I need to give a brief overview of models of duration perception.

1.2.1.2 Duration perception

There is no consensus regarding the mechanisms underlying (sub-second) duration perception (Grondin, 2010; Ivry & Schlerf, 2008). A rather consistent observation is that duration perception can be modulated by a multitude of non-timing related factors and thus is prone to illusions (Eagleman, 2008). However, widely differing models have been proposed for duration perception. Some assume a central, dedicated timer process that is often framed in a cognitivist 'box and arrow' style. An example of this

is Gibbon's 'Scalar Expectancy Theory' and its variations (Gibbon, Church, & Meck, 1984; Gibbon, 1977). In this popular model a 'pacemaker' emits regular pulses that that are sent to an 'accumulator' as long as a 'switch' is closed. The accumulated pulses are kept in a working memory space that can be compared to the number of pulses of a reference memory. Finally this comparison leads to a decision about duration comparisons. Proponents of this kind of model have tried to identify an internal 'pacemaker' as either the cerebellum, the basal ganglia, prefrontal or parietal cortex or a distributed system involving these and other structures (Buhusi & Meck, 2005; Coull, Cheng, & Meck, 2011; Grondin, 2010; Ivry & Schlerf, 2008). On the other hand different models propose that duration perception relies on 'intrinsic' properties of sensory neural processes (Ivry & Schlerf, 2008). For instance, (Karmarkar & Buonomano, 2007) showed that a simulated neural network with short-term plasticity shows distinct patterns of activity depending on stimulus duration and inter stimulus intervals. This is due to the state dependent nature of the network – the recent stimulus history will influence the response of the network to a present stimulus. Based on these simulated properties they predicted that duration judgments of human observers would depend on the recent stimulus history, which they could confirm empirically. Other models are simpler, in just assuming that the amplitude of neural activity integrated over time in a given sensory area serves as an estimator for stimulus duration (Grondin, 2010; Ivry & Schlerf, 2008). Evidence for this view comes from the correlation between sensory specific neural amplitude and duration estimates. For instance, visual adaptation to a flickering stimulus leads to shorter duration perception for stimuli

presented at the same visual field location (Bruno, Ayhan, & Johnston, 2010; Johnston, Arnold, & Nishida, 2006).

1.2.1.3 Cross-modal effects on duration perception

Several studies describe cross-modal effects on subjective duration perception (Chen & Yeh, 2009; Donovan et al., 2004; Klink et al., 2011; Walker & Scott, 1981). The first finding in this regard is that perceived auditory stimulus durations are expanded relative to perceived visual durations, and that the perceived duration of audiovisual stimuli is more similar to the one for auditory stimuli (Walker & Scott, 1981). As mentioned above this kind of auditory dominance is thought to reflect higher reliability of the auditory system for temporal judgments (Chen & Yeh, 2009). In line with, and extending, this hypothesis is the observation that temporal auditory dominance can be reversed for a high ratio of visual to auditory stimulus reliability (Burr, Banks, & Morrone, 2009; Wada, Kitagawa, & Noguchi, 2003). Most authors interpret their findings in the broad framework of Gibbon's scalar expectancy theory (Gibbon et al., 1984). For instance, the relative expansion of perceived auditory durations has been interpreted to reflect a faster auditory 'pace maker' (Chen & Yeh, 2009). However, this interpretation cannot explain all findings of cross-modal modulation of duration perception. For instance, the perceived duration of visual flashes critically depends on the duration of co-occurring sounds (Donovan et al., 2004; Klink et al., 2011).

Donovan et al. (Donovan et al., 2004) investigated the influence of task-irrelevant auditory information on a visual task, with participants judging whether two sequential visual events were presented for the same

or different durations (with fixed flash durations of 50 and 150 ms). They found that co-occurring sounds could decrease performance on the task if they were incongruent with regard to stimulus duration. However, this was only true if the onset of sounds and flashes was synchronous and for these stimulus durations. If the longer flash was prolonged to 250 ms duration (and participants performed closer to ceiling) the effect was abolished. More recently, Klink et al. (Klink et al., 2011) replicated these findings and the 'auditory expansion' phenomenon conceptually. Additionally they showed that flashes (of 500 ms duration) accompanied by a sound were perceived as lasting shorter or longer than a unimodal flash of same duration, depending on the duration of the co-occurring sound. They interpreted this to reflect a ventriloquist-like capture of visual stimulus on- and offsets by sounds, which would translate to changes in the timing of 'mode switch closures' in the above mentioned model of an internal clock (Klink et al., 2011).

This leaves at least two questions open, that I will try to address in Chapter 3. The first question is whether congruent sounds can *improve* visual duration comparisons relative to judgments regarding a pair of unimodal flashes. This would not be predicted by the interpretations of Klink et al., but would fit with the general framework of optimal cue integration. The second question is whether co-occurring auditory stimuli can modulate the *actual* duration of the perception of short visual stimuli. This is an alternative possibility to the interpretation that co-occurring auditory stimuli modulate processes specific to duration judgments. Co-occurring stimuli could alter visual perception *itself* and the associated

neural activity. This would be quite different from a scenario in which auditory stimuli would just modulate processes of a centralized and supramodal 'clock', assumed to be distinct from the sensory events themselves.

1.2.1.4 Sound induced flash illusion

The latter scenario becomes plausible in light of the *sound induced flash illusion*. Shams et al. (Shams et al., 2000) found that illusory percepts of multiple flashes can be induced when a single flash is accompanied by a sequence of multiple beeps. The illusion goes along with modulation of activity in striate and extrastriate visual areas of the cortex (Watkins, Shams, Tanaka, Haynes, & Rees, 2006). The multimodal modulation of visual areas is at least partly illusion-specific and has a very early component (30-60ms after the second bleep), as well as a later component (around 130ms post bleep) (Mishra, Martinez, Sejnowski, & Hillyard, 2007; Shams, Iwaki, Chawla, & Bhattacharya, 2005). The authors discuss early modulation as resting on possibly direct communication between primary auditory and primary visual cortex ((Clavagnier, Falchier, & Kennedy, 2004; Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2003); also see below, 1.3). Further, the illusory flash shares features of the physical one. It even provides a psychophysically detectable benefit to "see" the flash again when judging visual features like the orientation of a flashing grating (Berger, Martelli, & Pelli, 2003). The sound induced flash illusion thus provides evidence for the modulation of visual perception by audition, relevant for the second question mentioned above. But the illusion will be of interest for its own sake in Chapter 4. Previous studies found that

participants vary with regard to the fraction of trials in which they perceive the illusion (Mishra et al., 2007; Watkins et al., 2006). Individual differences in proneness to the illusion correlate not only with the magnitude of illusion associated ERPs but also with supra-additional multisensory ERPs in trials with two bleeps and two flashes (that do not induce any illusion) (Mishra et al., 2007). This suggests that these inter-individual differences might reflect traits of the observers rather than measurement noise. Further, Shams and colleagues found the sound induced flash illusion to be in line with the theory of statistically optimal weighting of sensory channels (Ernst & Banks, 2002; Shams, Ma, et al., 2005). As mentioned above, this theory proposes that the weighting of sensory channels that together form an integrated percept depends on their relative levels of signal-to-noise ratio. In case of the sound induced flash illusion the number of events is easier to tell for auditory beeps than for visual flashes (Shams, Ma, et al., 2005). Thus, if individual differences in proneness to the sound induced flash illusion are observer traits, they might reflect differences in the relative weighting of modalities between observers (Giard & Peronnet, 1999).

In Chapter 4 I will test whether the finding of individual differences in proneness to the sound induced flash illusion can be replicated and whether these differences are reliable. Such individual differences could go along with individual differences in brain morphology that might elucidate neural mechanisms of the illusion (Kanai & Rees, 2011). I will thus test whether individual differences in proneness to the illusion are correlated with systematic differences in brain morphology (Ashburner, 2009).

1.2.2 Auditory modulation of neural activity in the visual system

1.2.2.1 Animal studies

The behavioral studies discussed so far clearly demonstrate that audiovisual perception involves integration of sensory information across senses. But how do our brains combine information from different sensory streams? The earliest stages of cortical sensory processing were long thought to be unimodal and multisensory processing to be restricted to dedicated convergence areas (Mesulam, 1998). However, during the past decade new anatomical and functional evidence for multisensory interactions were found even at the level of primary sensory areas (Driver & Noesselt, 2008; Klemen & Chambers, 2012).

Tracer studies provide anatomical evidence for multisensory interactions at early stages of cortical processing (here referred to as 'early multisensory interactions' for convenience, not necessarily implying temporal precedence). There are direct feedback connections from primary auditory and multisensory areas to V1 and V2 in macaque (Clavagnier et al., 2004; Falchier et al., 2002; Rockland & Ojima, 2003) and similar connections in rodents (Allman et al., 2008; Budinger, Heil, Hess, & Scheich, 2006). Although some bimodal neurons can be found even in primary sensory areas (i.e. neurons that can be driven by either visual or auditory input, e.g. (Fishman & Michael, 1973)), the effect of direct cross-modal connections seems to be modulatory, rather than driving. Recent evidence from cats and rodents points to subthreshold modulation of 'unimodal' visual neurons (that cannot be driven by auditory input alone) as the dominant form of multisensory interaction in early visual cortex (Allman & Meredith, 2007;

Allman et al., 2008; Allman, Keniston, & Meredith, 2009; Iurilli et al., 2012). Early multisensory interactions also result in phase resetting of ongoing oscillations, thereby modulating and aligning the periodic excitability of affected neurons (Lakatos et al., 2009; Lakatos, Chen, Connell, Mills, & Schroeder, 2007; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008).

1.2.2.2 Human studies

In humans, cross-modal interactions modulate the amplitude or can drive neural signals in early visual cortex, as indexed by Blood Oxygenation Level Dependent (BOLD) fMRI (e.g. (Macaluso, Frith, & Driver, 2000; Martuzzi et al., 2007; Meienbrock, Naumer, Doehrmann, Singer, & Muckli, 2007; Noesselt et al., 2007; Watkins et al., 2006)), event-related potentials (ERPs) (e.g. (Cappe, Thut, Romei, & Murray, 2010; Molholm et al., 2002)) and transcranial magnetic stimulation (TMS) excitability (e.g. (Romei, Murray, Cappe, & Thut, 2009)). Cross-modal phase reset of ongoing oscillations in visual cortex is found in human magnetoencephalography (MEG; (Luo, Liu, & Poeppel, 2010)) and electroencephalography (EEG) consistent with phase-locked periodic modulations of perceptual performance (Naue et al., 2011; Romei, Gross, & Thut, 2012; Thorne, De Vos, Viola, & Debener, 2011).

1.2.3 Naturalistic stimuli and stimulus encoding

The effects discussed so far are typically found for highly artificial stimuli, like short flashes of Gabor patches and click sounds or beeps. Also,

auditory enhancement of visual detection performance and brain activity in visual areas typically is greatest for weak, peri-liminal visual stimuli (Noesselt et al., 2010). This raises the question which role audiovisual integration plays for richer, ecologically valid stimuli. Two recent studies in monkey used naturalistic video-type stimuli and found a new type of cross-modal interaction: enhancement of stimulus information carried by spike trains (Dahl, Logothetis, & Kayser, 2010; Kayser, Logothetis, & Panzeri, 2010). When monkeys are presented with naturalistic sound stimuli, accompanying visual stimulation reduces the mean firing rate of primary auditory cortex neurons (Dahl et al., 2010; Kayser et al., 2010). Moreover, inter-trial variability of spike trains is greatly reduced, thus enhancing mutual information between stimuli and spiking patterns. This effect is significantly stronger when the auditory and the visual input are congruent (Kayser et al., 2010). Visual neurons in STS show a similar behaviour for naturalistic visual stimuli (Dahl et al., 2010). Their response amplitude is somewhat reduced for bimodal audio-visual stimulation and the stimulus information carried by spike patterns is affected by multisensory context: incongruent sounds significantly worsen stimulus decoding based on spike trains.

In Chapter 5 I will test whether multisensory modulation of stimulus encoding extends to humans and to early visual cortices (V1-3).

1.3 Perceptual load and the spatial tuning of visual neural population responses

1.3.1 Introduction and overview

According to William James, attention 'implies withdrawal from some things in order to deal effectively with others' (James, 1890). But what does this withdrawal entail? Load theory (Lavie, 1995, 2005) proposes that increasing processing load associated with an attended target will suppress perceptual processing of distractors. According to load theory overall processing capacity has a fixed limit. Moreover, it tends to 'spill over' to task irrelevant distractors as long as task associated perceptual load isn't exhausting the available capacity.

This proposition was inspired by the observation of systematic differences between studies supporting early (e.g. (Broadbent, 1952)) vs. late (e.g. (Eriksen & Eriksen, 1974)) attentional selection. In a nutshell, models of early attentional selection hold that selection takes place early enough to prevent or at least attenuate processing of distractors beyond a fundamental sensory level that is restricted to the analysis of basic physical stimulus properties (e.g. (Broadbent, 1957); c.f. (Treisman, 1969)). Late selection models, on the other hand, argue that attentional selection requires full processing of a stimulus – up to a semantic level and equivalent to the processing required for stimulus perception (e.g. (Deutsch & Deutsch, 1963)). Lavie and Tsal (Lavie & Tsal, 1994) found that studies supporting early selection usually involved stimuli placing higher perceptual load on participants than the studies supporting late selection models. 'Higher

perceptual load' in this case refers to e.g. the use of bigger set sizes in search displays or to the use of stimuli that renders the discrimination between targets and distractors harder like targets defined by a feature conjunction vs. a simple feature. The main prediction of load theory – that systematically varying perceptual load will affect perceptual distractor processing – has since been confirmed numerous times on both, the behavioural and neural level (Lavie, 2005).

The finding of reduced distractor processing under high perceptual load will be of central interest in the context of this thesis (and thus be reviewed in more detail below). The experiment I will present in Chapter 6 probed a candidate neural mechanism of distractor suppression under high perceptual load. Specifically, it investigated whether perceptual load affects the spatial tuning of visual cortex neural populations responding to task-irrelevant stimuli. The idea that perceptual load might affect distractor processing by modulating the spatial tuning of neural populations in visual cortex is inspired by studies showing effects along these lines for spatial attention (that will also be reviewed below; c.f. (Anton-Erxleben & Carrasco, 2013).

Although not of central interest here, I should briefly mention that load theory has been extended during the past decade. New findings revealed that *cognitive* load (induced by e.g. working memory tasks) affects distractor processing in the opposite way of *perceptual* load (i.e. it *in*creases distractor interference; e.g. (de Fockert, Rees, Frith, & Lavie, 2001; Lavie, Hirst, de Fockert, & Viding, 2004); c.f. (Lavie, 2005, 2010)). This has been integrated in the theory as putting demand on a process governing the

allocation of limited resources. Cognitive or working memory load in this framework hampers the process prioritizing attended over unattended stimuli (e.g. (Lavie et al., 2004)). Thus it is important to distinguish between *perceptual* and *cognitive* (or working memory) load. Whenever the term *load* is used without further qualification I will refer to the former.

1.3.2 Effects of perceptual load on distractor processing

1.3.2.1 Behavioral studies

As mentioned above, a central prediction of load theory is that high target-associated load will affect perceptual processing of task irrelevant stimuli. This prediction has been tested and supported in many behavioural and neurophysiological studies. In this section I will summarize and review some of these findings.

The earliest studies on the effects of perceptual load (e.g. (Lavie & Cox, 1997; Lavie, 1995) used response competition tasks to measure the interfering effect of potentially distracting flanker stimuli (c.f. (Eriksen & Eriksen, 1974)). Typically, participants are instructed to indicate as quickly as possible with a button press which of two target letters appeared at the centre of a display. A peripheral distractor letter will accompany the target letter at the centre and can be either congruent to the target letter, incongruent (i.e. the other target letter) or neutral (i.e. an entirely different letter). The typical finding in such response competition tasks is that incongruent flanker stimuli yield longer reaction times than neutral flankers that in turn yield longer reaction times than congruent flankers (thus

supporting the notion of late attentional selection, after analysis of at least lexical information).

Crucially, studies on perceptual load introduce an additional manipulation, namely of target-associated load. For instance, embedding the target in a central search set of non-target letters reduces or even abolishes the effect of peripheral distractors (e.g. (Lavie, 1995), c.f. **Figure 1-1**a) and d)). The critical role of the central search set seems to be the introduced perceptual load. If the search set is too small, or homogeneously different from the target and the target thus pops out the effect of peripheral distractors remains (Lavie & Cox, 1997). Other experiments show that the effects of peripheral distractors can be diminished even without changing the stimuli at all – target associated load can be manipulated by changing the definition of targets e.g. from simple features to target defining feature conjunctions (e.g. (Bahrami, Carmel, Walsh, Rees, & Lavie, 2008; Carmel, Thorne, Rees, & Lavie, 2011; Lavie, 1995); c.f. (Cartwright-Finch & Lavie, 2007)).

High target-associated load does not only reduce task-interference by distractors. A series of behavioural studies shows that almost all aspects of basic stimulus processing for distractors or task-irrelevant stimuli are affected by load. High load reduces attentional capture by task irrelevant, salient stimuli like cartoon characters ((Forster & Lavie, 2008); c.f. **Figure 1-1** b) and d)) and can induce inattentional blindness for less salient stimuli ((Cartwright-Finch & Lavie, 2007); c.f. **Figure 1-1** c) and e)). It can eliminate negative priming effects of peripheral distractors; a finding that suggests that high load induces early target selection rather than active

suppression of distractors (because negative priming is thought to reflect such suppression; (Lavie & Fox, 2000), but c.f. (Fitousi & Wenger, 2011) for a critique). High central load reduces detection sensitivity for peripheral targets in dual task paradigms ((Carmel et al., 2011; Russell, Malhotra, & Husain, 2004); c.f. (Plainis, Murray, & Chauhan, 2001; Zenger, Braun, & Koch, 2000)) and it interferes with object recognition in the periphery (as indicated by reduced positive priming; (Lavie, Lin, Zokaei, & Thoma, 2009)). High central load further reduces orientation specific adaptation to peripheral gratings (even if they are masked and not consciously perceived; (Bahrami et al., 2008; Bahrami, Lavie, & Rees, 2007)) and motion aftereffects for task irrelevant stimuli ((Rees, Frith, & Lavie, 1997); but see (Morgan, 2013, 2011, 2012) for a critique). Finally, high load can reduce the flicker-fusion threshold for a task-irrelevant foveal stimulus and is thus thought to affect the temporal resolution of the representation of task irrelevant stimuli (Carmel, Saker, Rees, & Lavie, 2007).

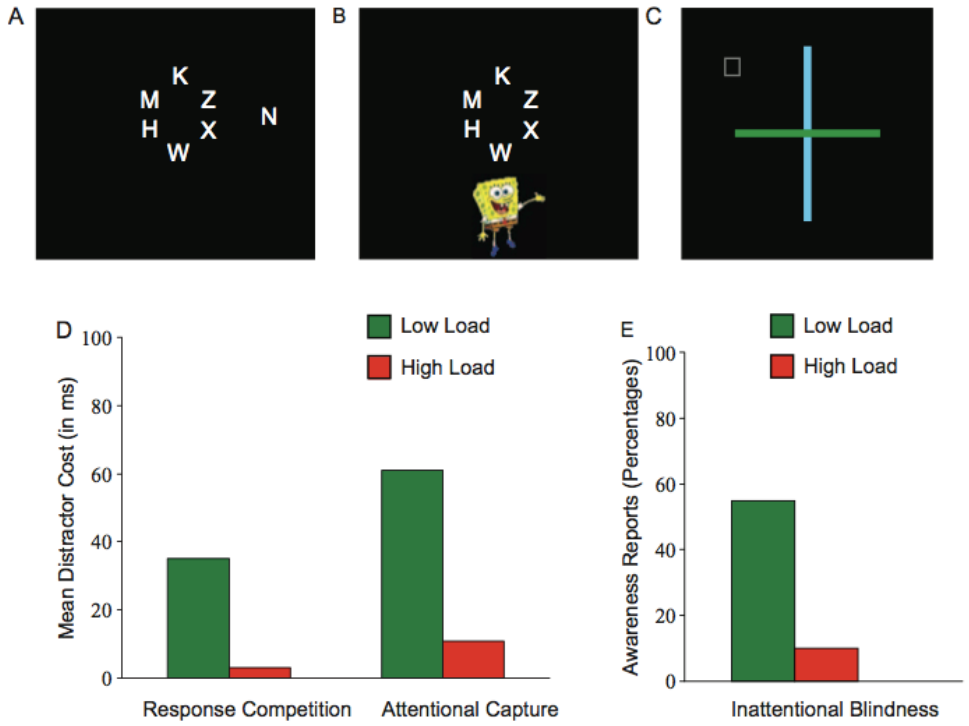

Figure 1-1 Example tasks and results for effects of perceptual load on distractor processing. **A** Example of a response competition task with high target-associated load (from (Forster & Lavie, 2008)). Participants had to indicate which of two-predefined letters (X or N) is present in the central ring of letters. The example trial depicted is incongruent (the flanker letter to the right is mismatching the central target). Perceptual load is induced by the array of letters surrounding the target making identification of the target harder. In the low load condition all letters but the target would be replaced by the latter 'o'. The left hand side of panel **D** shows the corresponding results: Distractor incongruence slowed participants down only in the low load condition. Equally, an entirely unrelated but salient stimuli (panel **B**) only slowed participants down in the low load condition (panel **D**). **C** Example of a stimulus that can be used in identical form for low and high perceptual load. Cartwright-Finch and Lavie (Cartwright-Finch & Lavie, 2007) asked participants to indicate either which of the arms of a cross was green (low load) or which one was the longer one (high load). They then presented an unexpected additional stimulus (small square in the top left quadrant) and asked participants after the respective trial whether they noticed the extra stimulus. Participants reported distractor awareness with much higher probability in the low load condition (**E**). Figure source: (Lavie, 2010).

1.3.2.2 Neurophysiological studies

Besides these behavioural findings, several neurophysiological studies have shown that the neural responses evoked by distractors are reduced by target-associated load. For instance, the threshold for TMS-induced motion phosphenes is increased under high perceptual load (i.e. the excitability of human MT is reduced; (Muggleton, Lamb, Walsh, & Lavie, 2008)). High target-associated load further reduces distractor related early event related potentials over occipital cortex ((Handy, Soltani, & Mangun, 2001; Rauss, Pourtois, Vuilleumier, & Schwartz, 2009; Rorden, Guerrini, Swainson, Lazzeri, & Baylis, 2008); c.f. (O'Connell, Schneider, Hester, Mattingley, & Bellgrove, 2011; Rauss, Pourtois, Vuilleumier, & Schwartz, 2012)). In monkeys, attention outside a neurons receptive field will yield greater suppression of this neuron's responses if the attended task is harder (Chen et al., 2008). Finally, a number of studies investigated distractor-related BOLD responses in visual cortex. High load reduces the BOLD response in human MT, the superior colliculi and early visual cortex to radial motion distractors (Rees et al., 1997). It reduces BOLD responses to high contrast checkerboard stimuli in LGN (O'Connor, Fukui, Pinsk, & Kastner, 2002) and V1-V4 (e.g. (Schwartz et al., 2005)). The latter effect diminishes from about 8 degrees eccentricity (Schwartz et al., 2005), is excessively strong in the right hemisphere for parietal neglect patients (Vuilleumier et al., 2008) and weakened in individuals with autism spectrum conditions (Ohta et al., 2012). Furthermore, even BOLD responses evoked by masked, invisible distractor stimuli are reduced under high load in V1 (Bahrami et al., 2007). Finally, high load reduces BOLD responses to

colourful distractor stimuli in human V4 (Pinsk, Doniger, & Kastner, 2004), an effect that is reduced in participants suffering from major depression (Desseilles et al., 2009).

Taken together, there is good evidence that high target associated load reduces distractor related task-interference, sensitivity, adaptation and neural responses. However, a strand of work on the effects of spatial attention suggests that attention cannot only modulate the amplitude of neural responses. It can also influence the spatial preference or tuning properties of visual neurons. In Chapter 6 I will test whether such effects can be observed for perceptual load as well. Therefore I will review previous findings in the context of spatial attention below.

1.3.3 Spatial attention, neural resolution and acuity

1.3.3.1 The classic study by Moran & Desimone (1985)

In a classic study Moran & Desimone (Moran & Desimone, 1985) presented monkeys with two stimuli that were both placed within the receptive field of a recorded V4 neuron. The stimuli were purposefully chosen: One stimulus was effective in driving the neuron's response while the other one was ineffective (e.g. due to the neuron's color preferences). Crucially, the monkey then had to solve a match to sample task involving one of the stimuli. This required the monkey to maintain fixation but to covertly shift its attention to the respective stimulus. The measure of interest was the spike rate of the neuron depending on which of the two stimuli was attended. The majority of cells recorded showed more vigorous

firing (i.e. about twice as often) when the monkey attended the effective vs. ineffective stimulus.

Similar results were observed for neurons in V2 (Luck, Chelazzi, Hillyard, & Desimone, 1997; Reynolds, Chelazzi, & Desimone, 1999). The response evoked by an attended stimulus (out of two in the receptive field) is similar to the response evoked by this stimulus in isolation (Reynolds et al., 1999). Furthermore, the attentional gain can reflect an effective gain in stimulus contrast rather than a mere multiplicative gain of response amplitude ((Reynolds, Pasternak, & Desimone, 2000); but c.f. (Reynolds & Heeger, 2009)). Finally, higher task demands leading to an increased attentional load enhance the observed effect (Spitzer, Desimone, & Moran, 1988).

From the outset this result was interpreted to reflect changes in the receptive field properties of the neuron. If attention moved the center of the receptive field towards the attended stimulus, or shrank the receptive field around the attended stimulus, this would yield the observed pattern of results (Moran & Desimone, 1985).

1.3.3.2 Receptive field changes in single neurons

The hypothesis that the spatial preference of visual neurons can be modulated by spatial attention has since been confirmed in a series of experiments (for an overview see (Anton-Erxleben & Carrasco, 2013)). For instance neurons in macaque V4 shift their peak sensitivity towards the attended location (even if it is outside its classical receptive field; (Connor, Gallant, Preddie, & Van Essen, 1996; Connor, Preddie, Gallant, & Van Essen,

1997)). Mapping the receptive fields of neurons in LIP during free gaze vs. attentive fixation reveals smaller and more well defined receptive fields during fixation (Ben Hamed, Duhamel, Bremmer, & Graf, 2002). In a study very close to the original design of (Moran & Desimone, 1985), Womelsdorf et al. (Womelsdorf et al., 2006; Womelsdorf, Anton-Erxleben, & Treue, 2008)manipulated which of two (ineffective) stimuli were attended while recording from an MT neuron. At the same time mapping stimuli were used to probe and reconstruct the receptive field of the MT neuron in a dense fashion. Receptive fields shifted towards the attended stimulus (even if it was in the opposite hemifield) and if the attended stimulus fell within the neurons receptive field it shrunk around the stimulus. This shift of the receptive field applies to both, its center and its suppressive surround and the suppressive surround of the receptive field becomes deeper or shallower depending on whether the attended target falls within or outside its classical receptive field, respectively (Anton-Erxleben, Stephan, & Treue, 2009). Finally, when traveling mapping stimuli are attended receptive field sizes in area MT expand, i.e. they contain the stimulus for longer (Niebergall, Khayat, Treue, & Martinez-Trujillo, 2011).

1.3.3.3 Neurophysiological studies in humans

In humans spatial attention has been shown to be accompanied by retinotopically specific enhancement of BOLD responses many times probing both endogenous (e.g. (Bouvier & Engel, 2011; Brefczynski & DeYoe, 1999; Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; Martínez et al., 1999; Morawetz, Holz, Baudewig, Treue, & Dechent, 2007;

Muller, Bartelt, Donner, Villringer, & Brandt, 2003; Tootell et al., 1998) and exogenous (e.g. (Liu, Pestilli, & Carrasco, 2005)) cues to manipulate spatial attention. Enhancement of neural responses at attended locations is associated with a suppression of responses to stimuli in the area surrounding the attended location as indicated by high density MEG (Hopf et al., 2006; Hopf, Boehler, Schoenfeld, Heinze, & Tsotsos, 2010) and fMRI (Heinemann, Kleinschmidt, & Müller, 2009; Müller & Kleinschmidt, 2004).

Of greater direct relevance in the present context is the observation that the position tuning of BOLD responses in human V1-4 changes with spatial attention (Fischer & Whitney, 2009). When participants attend an array of Gabor patches in the periphery the elicited patterns of BOLD responses will be more distinct for different eccentricities when these patches are attended vs. attention being paid to a central fixation task. In line with this finding is the observation that retinotopic mapping results become more reliable when the mapping stimulus is attended (Bressler & Silver, 2010). Finally, the finding that a cholinesterase inhibitor decreases the spread of visual BOLD responses has also been interpreted to support the notion of spatial attention refining the spatial selectivity of neural responses (Silver, Shenhav, & D'Esposito, 2008).

1.3.3.4 Behavioral studies in humans

For humans the hypothesis of spatial attention modulating representational resolution has been directly tested and supported on the behavioral level. For example, both endogenous and exogenous cues can modulate Landolt acuity (Montagna, Pestilli, & Carrasco, 2009). Valid cues,

yielding a covert shift of attention towards the location at which a Landolt stimulus is flashed result in enhanced perceptual acuity (i.e. an enhanced ability to discern stimuli) relative to a neutral, uninformative cue. Invalid cues that direct attention away from the location at which the Landolt stimulus will be shown have the opposite effect. Similar results have been found for Vernier acuity ((Shiu & Pashler, 1995); c.f. (Golla, Ignashchenkova, Haarmeier, & Thier, 2004; Shalev & Tsal, 2002)).

A second line of studies used texture segmentation tasks to probe changes in the effective visual resolution of human observers. Texture segmentation tasks typically require observers to detect the presence of a target patch within a background texture. For instance the background texture might consist of lines with a fixed length, spacing and orientation while the target patch contains lines of a different orientation (e.g. (Kehrer, 1987)). When this sort of stimulus is presented only very briefly, observers show a 'central performance drop', i.e. they show better sensitivity for the target when it is presented in a mid-range of eccentricities and become worse for both more eccentric and more central target locations (e.g. (Kehrer, 1987, 1989; Potechin & Gurnsey, 2003)). Crucially, the optimal eccentricity to detect the target patch varies with its size (Gurnsey, Pearson, & Day, 1996; Kehrer, 1989). The effect is thus thought to reflect the eccentricity dependent visual resolution of the image representation (Gurnsey et al., 1996; Kehrer, 1997). This opens up the possibility to test the effect of attention on the effective spatial resolution of the observer. In a crucial study Yeshurun & Carrasco (Yeshurun & Carrasco, 1998) asked participants to indicate in which of two intervals a briefly presented

background texture contained a target patch. Both intervals were preceded by a brief cue drawing covert spatial attention to the target location in the target interval and to a separate location in the other interval. As expected this increased performance relative to a neutral (i.e. spatially non-informative) cue for peripheral locations. Crucially, however, performance *de*creased at foveal locations. That is, the central performance drop became worse and peak performance was shifted outwards. Further, the range of inner eccentricities for which performance was worsened by attention scaled with stimulus size. This is in line with the idea that exogenous attention increases perceptual resolution of the image. At foveal locations the already high spatial resolution would shift towards a spatial frequency preference too high for the target size and thus away from the optimum, while the opposite would be true for more peripheral locations (c.f. (Anton-Erxleben & Carrasco, 2013; Carrasco, Loula, & Ho, 2006)).

This finding has since been replicated multiple times (e.g. (Talgar & Carrasco, 2002; Yeshurun & Carrasco, 2000)). It is complemented by another line of research concerning changes of *appearance* induced by attention. Results showing that exogenous attention increases the perceived extent of circular visual stimuli (Anton-Erxleben, Henrich, & Treue, 2007) could be interpreted as reflecting increased spatial resolution. Another interpretation would be perceptual repulsion - if attention directed towards the center of the stimulus attracts receptive fields a subjective compression of space and thus repulsion of its edges might be the consequence. The latter interpretation seems to be supported by results showing that the contours of an oval are subjectively repelled by exogenous attention, rendering the

oval taller or wider depending on the cued locations (Fortenbaugh, Prinzmetal, & Robertson, 2011).

Endogenous spatial attention enhances texture segmentation performance at all eccentricities – i.e. there is no performance decrease at foveal locations (Yeshurun, Montagna, & Carrasco, 2008). This is interpreted to reflect a more flexible nature of endogenous attention as a means of allocation of resources. However endogenous attention can increase the *perceived* spatial frequency of Gabor patches flashing up at an attended location (Abrams, Barbot, & Carrasco, 2010).

Taken together there is good behavioral and neurophysiological evidence that endogenous and exogenous spatial attention can modulate the spatial tuning properties of visual representations. However, most neurophysiological evidence stems from monkey studies investigating endogenous attention while most studies in humans concentrated on behavioral effects of exogenous attention. In Chapter 6 I will ask whether similar effects can be observed for perceptual load and population responses in human visual areas V1-3. In other words, whether perceptual load can modulate neural preferences for the feature of location. In the final chapter I will probe a somewhat different question – namely whether there are interactions between (neural and perceptual) preferences for location and face features.

1.4 Face recognition (and canonical retinotopic locations of face features)

1.4.1 Introduction and overview

The experiment presented in Chapter 7 addresses the question of whether canonical retinotopic stimulus locations matter for the recognition of face features. Here, I aim to give a brief introduction to the main concepts of the broader face recognition literature. Specifically, I will introduce two popular models of face recognition, some effects pointing to the importance of configural aspects of faces, main findings regarding the neural underpinnings of face processing in humans and monkeys and studies investigating eye movements towards faces. The introduction and discussion sections of Chapter 7 will build upon these concepts to discuss location specificity in face processing more specifically. For a succinct general introduction to face recognition also see (Bate, 2013).

1.4.2 Cognitive models of face recognition

1.4.2.1 The Bruce & Young (1986) model

Bruce and Young (Bruce & Young, 1986) proposed what became probably the most influential model of face recognition to date. Their model first aims to distinguish and describe different types of information or 'codes' that observers extract from a face. It then proposes a functional model of this extraction process.

According to this model, observers extract seven types of information from a face that are labelled *pictorial*, *structural*, *visually derived semantic*,

identity-specific semantic, *name*, *expression* and *facial speech* codes. *Pictorial* information refers to aspects of a face image that are picture-specific, for instance the lighting conditions of a photograph. *Structural* information relates to face features and their configuration in a manner that allows to individuate identity. In the context of this thesis it is interesting to note that Bruce and Young (Bruce & Young, 1986) make specific suggestions regarding the view (in)dependent nature of structural representations. They point out that

'[…] the range of transformations of viewpoint across which we need to recognize faces in everyday life is considerably smaller than the range of transformations involved in object recognition, so that it is conceivable that object-centred descriptions are less important to face recognition. People usually stand with their heads more or less upright, and indeed face recognition is particularly prone to disruption when faces are inverted […]. In addition, people will often look toward you, though recognition of profiles is of course quite possible. '

They go on to cite earlier findings of cells in monkey IT that are tuned to the view of faces with a specific angle of head rotation (e.g. (Perrett et al., 1985; Perrett, Rolls, & Caan, 1982); c.f. below) and conclude

'(*W)e propose that a familiar face is represented via an interlinked set of expression-independent structural codes for distinct head angles, with some*

codes reflecting the global configuration at each angle and others representing particular distinctive features.'

Visually derived semantic information refers to (relatively) invariant attributes of a face's owner, like age or sex, which can be inferred from an unfamiliar face. *Identity-specific semantic information* on the other hand builds upon memories about a familiar face owner, like the context in which they are usually encountered. *Names* are explicitly separated from other semantic information because retrieval and forgetting of names seems to be relatively independent of other biographical information in healthy observers and patients with recognition problems (e.g. (McWeeny, Young, Hay, & Ellis, 1987; Young, Hay, & Ellis, 1985)). Finally, *facial speech* and *expression* cues describe dynamic aspects of face features and their configuration, respectively.

The functional model of (Bruce & Young, 1986) proposes a sequence of steps to extract these different types of information (**Figure 1-2**). According to the model, the first step of face processing is *structural encoding*. This includes a set of view-centered descriptions that are used in expression and facial speech analysis as well as a transformation to more abstract, expression independent descriptions. Different aspects of the face can be emphasized in an attention dependent manner, labeled *directed visual processing*. In the context of this thesis *structural encoding* is of primary interest. However, the model by (Bruce & Young, 1986) emphasizes further, memory related steps that lead to recognition of familiar faces and

these will be summarized below.

The model holds that more abstract structural descriptions of a face are passed on to *Face Recognition Units* (FRUs) where they are compared to stored structural descriptions of faces. This, in turn, will result in a signal to the cognitive system indicating recognition and the strength of this signal will depend on the degree of resemblance between the encoded structure of the face and a stored exemplar. Furthermore, the output of the FRU is sent to *Person Identity Nodes* (PINs) that associate semantic information about familiar faces. PINs can also aid recognition via their output to the cognitive system (e.g. via priming, an idea that received more attention in a later version of the model; (Burton, Bruce, & Johnston, 1990)). The final stage of processing is retrieval of the name associated with a familiar face.

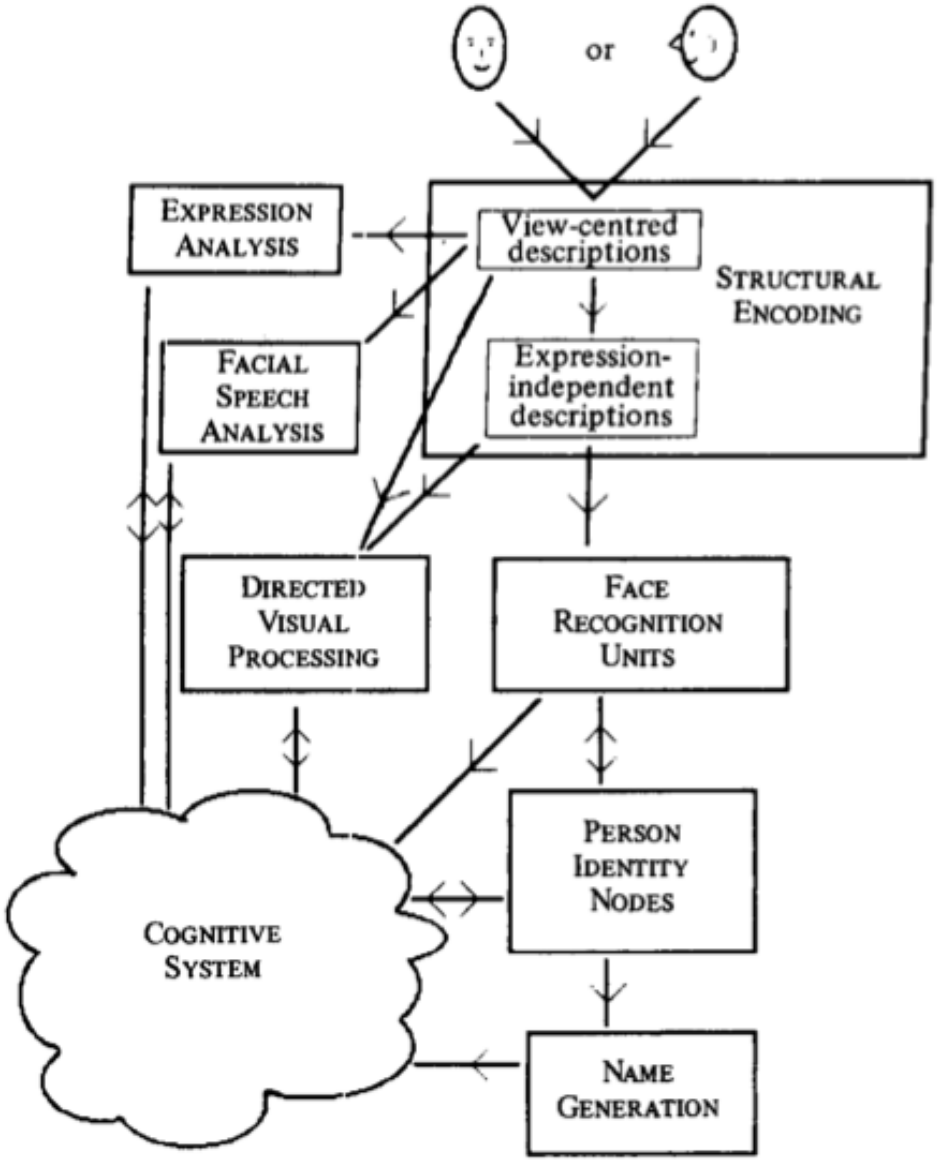

Figure 1-2 Bruce & Young's functional model of face recognition. See text for details and explanations. From (Bruce & Young, 1986)

1.4.2.2 Valentine's (1991) multidimensional face space

Valentine (Valentine, 1991) proposes that structurally encoded faces are represented in a way that is best described as a vector in a multidimensional 'face space'. The dimensions of this space would correspond to features enabling face discrimination, such as face shape or age, but are not specified. (Note that 'features' in this context refer to aspects of a face that can be configural, rather than to isolated face parts as in other contexts above and below). The origin of face space is assumed to reflect the mean value of previously experienced faces for each feature dimension. Further, faces are assumed to vary normally around this origin in each dimension (i.e. more extreme deviations from the mean of a given feature dimension occur less often). Similarity between faces would then be established as their distance[1] in this multidimensional face space.

[1] Note that (Valentine, 1991) distinguishes between norm and exemplar based versions of this model. In the norm-based model the origin of face space is an explicit representation of an average face or prototype, while in the exemplar based version it 'plays no part in encoding stimuli'. In the exemplar-based version of the model the origin 'merely indicates the point of maximum exemplar density' and thus 'it is more appropriate to consider faces as being encoded as points rather than vectors'. Mathematically this is equivalent, of course, and it is somewhat difficult to understand what Valentine is aiming to suggest. One more concrete way in which (Valentine, 1991) differentiates these different flavours of the model is the use of different face similarity measures for each. Exemplar based versions would imply face similarity to be encoded as Euclidean distance between points while a norm based version would suggest other measures of similarity „such as the dot product between two vectors". Note that the latter would have rather bizarre consequences. The dot product is defined as the product of vector magnitudes and the cosine of their angle. This would imply the similarity of faces with orthogonal vectors would always be zero, regardless of their distance in face space. Also, the similarity between faces and their caricature would *increase* as the caricatures become more extreme (because faces and their caricatures would be represented by collinear vectors).

This model potentially explains ease of recognition for distinct faces. According to (Valentine, 1991) more distinct faces would inhabit a less densely populated hypervolume of face space and would thus be more easily distinguished from 'false positive' candidate faces in memory. From this he derives the prediction that more distinct faces should be more tolerant to imprecise encoding (i.e. a greater level of uncertainty with regard to their exact location in face space). Indeed, face inversion (a manipulation assumed to affect encoding precision, c.f. below) affects recognition performance for distinct faces less than for typical faces. Finally, (Valentine, 1991) proposes that the idea of face space explains lower recognition performance for faces of other race (than the observer). Assuming less exposure to such faces the observer's face space would span dimensions that are more appropriate for own-race faces and consequentially lead to a dense clustering of other-race faces at a distance from the origin. Again, face inversion is predicted and found to interact with this effect. Face inversion has a more detrimental effect on recognition of other-race faces. This was predicted because less precise encoding leads to face confusion more easily for faces that populate a denser hypervolume of face space.

Later, this model gained additional support by the finding of face adaptation effects. Adapting to the distorted (morphed) version of a face causes the original image to appear distorted in the opposite way (Webster & MacLin, 1999). Furthermore, observers can adapt to naturally salient features of faces like gender or race (Webster, Kaping, Mizokami, & Duhamel, 2004) and identity (Leopold, O'Toole, Vetter, & Blanz, 2001; Rhodes & Jeffery, 2006). The latter type of study derived stimuli from a large

collection of three-dimensional face scans (n=200) in which each face was represented as a vector in a high-dimensional space encoding face shape and texture (with feature dimensions corresponding to location and colour values for each vertex of the sampled face surface; (Blanz & Vetter, 1999)). This allows the derivation of an average face across all faces in the database and to computationally determine a face's distance from the average. Crucially, it also allows interpolating faces along the axis spanning from an exemplar face to the average face and beyond, where they become 'anti' or 'opposite' faces (Leopold et al., 2001). When adapting to an 'anti-face' observers are more likely to (correctly or incorrectly) identify a test face as the corresponding original (Leopold, Bondar, & Giese, 2006). Finally, and of particular interest here, early reports of face adaptation effects emphasized their robustness towards minor retinotopic translations (e.g. (Leopold et al., 2006; Rhodes, Jeffery, Watson, Clifford, & Nakayama, 2003)) but later studies found that these effects are retinotopically confined to some extent (Afraz & Cavanagh, 2009, 2008).

1.4.3 Effects highlighting the importance of configural face processing

The models discussed above somewhat sidestep the question of *which* aspects of a face are encoded and used for recognition. As previously mentioned (Bruce & Young, 1986) suggest that both properties of face features (like eye colour) and their configuration (like inter-eye distance) are used for face recognition. This assumption has been confirmed

empirically (e.g. (Rhodes, 1988)) and several findings highlight the particular importance of configural processing for face recognition.

Face recognition performance is severely affected by inversion of face images – and significantly more so than recognition performance for other object categories (Schwaninger, Carbon, & Leder, 2003; Valentine, 1988; Yin, 1969). Studies varying configural aspects and those related to isolated features independently found that face inversion predominantly affects the recognition of configural aspects (Leder & Bruce, 2000; Schwaninger & Mast, 2005), specifically the accurate perception of vertical distances between features (Goffaux & Rossion, 2007).

When participants are asked to recognise the upper or lower half of a face their performance is detrimentally affected by the presence of a mismatching (but task irrelevant) face half (the *composite effect*: (Rossion, 2013)). This is true for the recognition of familiar faces (e.g. (Young, Hellawell, & Hay, 1987)) as well as for match-to-sample tasks using images of novel faces (e.g. (Laguesse & Rossion, 2013)), but only if the face halves are aligned exactly (Laguesse & Rossion, 2013). The effect is also smaller for inverted faces (Susilo, Rezlescu, & Duchaine, 2013). The *composite effect* has been interpreted as evidence for a mandatory holistic processing of faces (e.g. (Maurer, Grand, & Mondloch, 2002)). Observers seem unable to ignore irrelevant parts of a face even if this is disadvantageous for the task at hand.

The *part-whole* effect (Tanaka & Farah, 1993) also provides evidence for the importance of context in face processing. It describes a recognition advantage for face features when they are embedded in a face context. When observers learn the identity of faces they are significantly better at

discriminating between the target face and a version of this face in which one feature (like the nose) has been replaced than at discriminating between the isolated target and distractor features. However, this effect is only found for intact (unscrambled), upright faces and not for other object categories like houses (Tanaka & Farah, 1993).

Finally, (Maurer et al., 2002) propose a differentiation of 'configural face processing' into more clearly defined components. They refer to *first-order relations* as the basic configuration of any face (eyes above nose above mouth) that would enable recognition of faces as such. Sensitivity to *second-order relations* would describe sensitivity to distances among features and *holistic processing* the integration of those parts into a whole or 'Gestalt'.

1.4.4 Neural underpinnings of face processing

The functionalist models of face processing presented above do not explicitly address questions of neural implementation (e.g. (Bruce & Young, 1986)). Here, I turn to findings regarding the neural underpinnings of face processing.

Studies probing single neurons in macaque temporal cortex found cells with a preference for complex shapes, including hands and faces (e.g. (Gross, Rocha-Miranda, & Bender, 1972)). Neurons with a preference for face stimuli were found in the superior temporal sulcus (STS; e.g. (Perrett et al., 1985, 1982)) and inferior temporal cortex (IT; e.g. (Desimone, Albright, Gross, & Bruce, 1984; Gross et al., 1972)). Neurons in STS were found to be especially sensitive to gaze direction and expression (e.g. (Perret et al.,

1990)) while neurons with identity preferences appear to be more common in IT (e.g. (Hasselmo, Rolls, & Baylis, 1989)).

Generally, face sensitive cells vary with regard to their exact tuning properties. For instance, some cells prefer particular angles of head rotation (e.g. (Desimone et al., 1984; Perrett et al., 1982)). Some neurons are tuned to whole faces while others prefer face parts or isolated features (e.g. (Issa & DiCarlo, 2012; Perrett et al., 1982)) or show tuning for the spatial arrangement of features (Freiwald, Tsao, & Livingstone, 2009). Early studies emphasized the broad spatial tuning of face sensitive cells with large receptive fields (e.g. (Tovee, Rolls, & Azzopardi, 1994)), while later studies qualified this (e.g. (Op De Beeck & Vogels, 2000); see Chapter 7 for a more detailed discussion of this point).

fMRI guided single cell recordings in macaque revealed that face sensitive neurons are concentrated in patches along IT and STS in macaque (e.g. (Tsao, Freiwald, Knutsen, Mandeville, & Tootell, 2003; Tsao, Freiwald, Tootell, & Livingstone, 2006)). Up to 97% of sampled neurons in these *face patches* were found to be face sensitive (Tsao et al., 2006). Furthermore, face patches appear to show a functional gradient with a posterior-anterior progression from identity-invariant viewpoint preferences to viewpoint-invariant identity preferences (Freiwald & Tsao, 2010). Notably, the most posterior face patch (possibly a homologue to human OFA) contains cells preferring isolated eye stimuli and with receptive fields in the contralateral upper quadrant of the visual field (Issa & DiCarlo, 2012).

Generally, face sensitive patches of cortex (at least those in IT) have been interpreted as part of a 'ventral stream' for visual recognition (as

opposed to a dorsal stream for action related vision; (Goodale & Milner, 1992)). The relatively simple tuning properties of early visual cortex and more abstract tuning properties of IT neurons have been interpreted to reflect a general hierarchy along the ventral stream, from view specific descriptions to view-invariant identity representations (e.g. (Riesenhuber & Poggio, 1999), but see e.g. (Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013)).

In humans, fMRI studies have revealed a similar network of cortical patches that show a preference for faces and face parts above other object categories. Probably the most studied of these areas is the (right) fusiform face area (*FFA*; (Kanwisher, McDermott, & Chun, 1997; Kanwisher & Yovel, 2006)). Notably, activation of this area correlates with face identification performance across trials (Grill-Spector, Knouf, & Kanwisher, 2004) and is diminished by face inversion (e.g. (James, Arcurio, & Gold, 2013; Yovel & Kanwisher, 2005)). Stimulating neurons in this area results in perceptual 'metamorphosis' of faces (and only faces, (Parvizi et al., 2012)). Recently, Weiner and Grill-Spector proposed that FFA might be divided into two components, the posterior and mid fusiform face area (pFus and mFus; (Weiner & Grill-Spector, 2012, 2013)). These are separable based on their anatomy and cytoarchitecture (Weiner et al., 2014).

A more posterior, occipital face area has been found to overlap the inferior occipital gyrus (*IOG* or *OFA*: e.g. (Gauthier et al., 2000; Puce, Allison, Asgari, Gore, & McCarthy, 1996)). Patients with acquired face recognition deficits (prosopagnosia) mostly have lesions in the vicinity of this area (Bouvier & Engel, 2006). OFA is responsive to isolated face parts (Nichols,

Betts, & Wilson, 2010) and disruption of OFA using TMS particularly interferes with face recognition based on face parts (rather than their configuration; (Pitcher, Duchaine, Walsh, Yovel, & Kanwisher, 2011; Pitcher, Walsh, Yovel, & Duchaine, 2007).

Like monkeys, humans also have a face sensitive patch in posterior superior temporal sulcus (pSTS; e.g. (Hoffman & Haxby, 2000; Puce et al., 1996)). (Hoffman & Haxby, 2000) found that attention to gaze direction and face identity go along with increased response amplitudes in pSTS and FFA, respectively. This observation resonates with the functional differentiation of STS and IT neurons in macaque (see above). It led (Haxby, Hoffman, & Gobbini, 2000) to propose a rough mapping of Bruce & Young's (1986) functional model of face recognition onto face sensitive areas (**Figure 1-3**). According to this neural model, structural encoding of faces begins in IOG

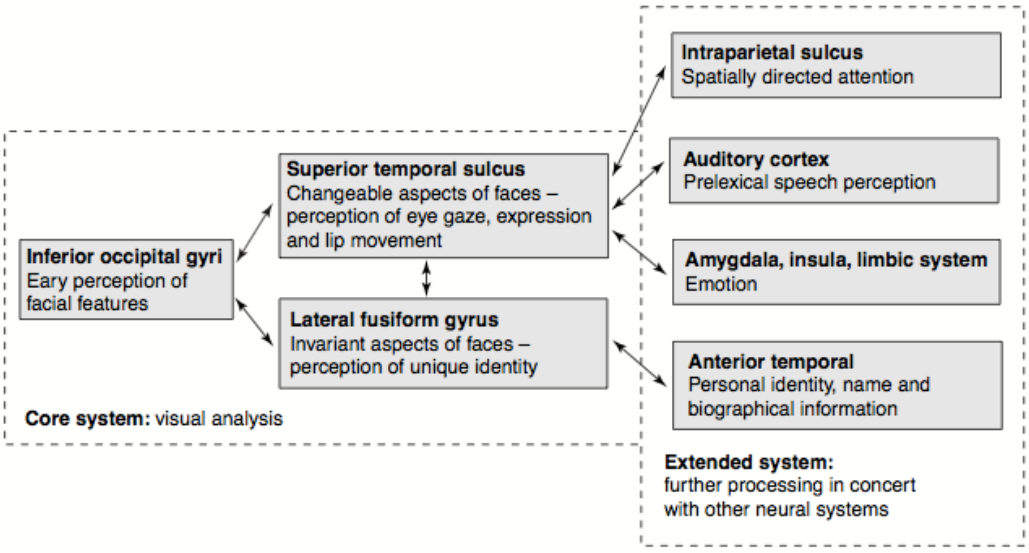

Figure 1-3 Haxby's model of face processing in face sensitive areas of visual cortex. C.f. Figure 1-2; See text for details and explanations. From (Haxby et al., 2000)

(or OFA) from where information is passed to STS and FFA. STS would analyse this information with respect to (potentially) dynamic and changeable aspects, like gaze direction while FFA would analyse identity specific invariant aspects of the face. Furthermore, these core regions of visual face analysis would engage in bidirectional communication with other regions of cortex in an extended face processing network (c.f. (Bruce & Young, 1986)). For instance, communication with auditory regions would aid speech perception, interaction with limbic structures would aid emotion recognition and anterior temporal cortex would be involved in identity matching and retrieval. Finally, bidirectional communication between STS and intra-parietal sulcus would enable spatial attention to modulate the process. It is worth noting that anterior temporal lobe indeed has been found to be associated with face memory (e.g. (Von Der Heide, Skipper, & Olson, 2013)) and patterns of activity in the anterior part of IT are identity specific (Kriegeskorte, Formisano, Sorger, & Goebel, 2007).

Finally, EEG and MEG studies found several face sensitive response components. The most prominent of these is the N170 (or 'M170' for MEG; e.g. (Bentin, Allison, Puce, Perez, & McCarthy, 1996)). It is strongest over right occipitotemporal cortex and has been found to be independent of face familiarity but delayed for inverted faces and thus interpreted to reflect the structural encoding stage of face processing (e.g. (Eimer, 2011)).

1.4.5 Eye movements towards faces

Retinotopic location(s) of faces and face features will crucially depend on eye movements towards faces. Human observers show a

stereotypical pattern of gaze behavior towards face stimuli that implies retinotopic heterogeneity for the frequency with which different face features will appear in the visual field.

The first fixations towards a face typically land on the central upper nose region, just below the eyes (Hsiao & Cottrell, 2008). In Western observers subsequent fixations are concentrated on a triangular region spanning from eyes to mouth (e.g. (Henderson, Williams, & Falk, 2005; van Belle, Ramon, Lefèvre, & Rossion, 2010; Walker-Smith, Gale, & Findlay, 1977)), while the pattern of fixations remains more concentrated for East Asian observers (Blais, Jack, Scheepers, Fiset, & Caldara, 2008; Miellet, Vizioli, He, Zhou, & Caldara, 2013). Interestingly the first two fixations seem sufficient for face recognition and no additional benefit is gained from subsequent fixations (Hsiao & Cottrell, 2008). Furthermore, fixations just below the eyes are optimal for face recognition performance, both in humans and according to an ideal observer model incorporating the foveated nature of the visual system (Peterson & Eckstein, 2012). Overall, observers tend to look at inner face features and avoid looking directly at outer features like the chin and upper forehead (cf**. Figure 1-4**).

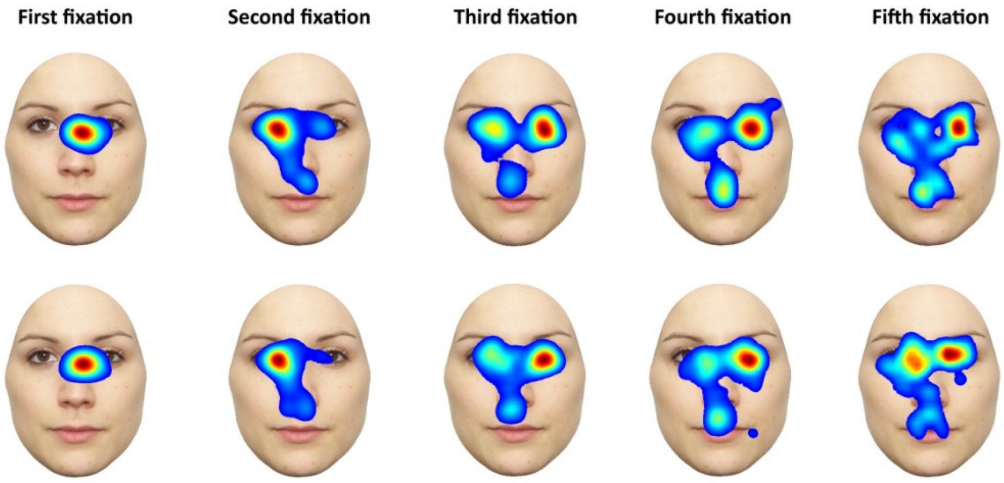

Figure 1-4 Typical gaze behavior towards faces. Heatmaps indicate the (relative) probability of fixations for different locations on an example face. Columns show heatmaps for the first to fifth fixation (from left to right) and the upper and lower row show data for personally familiar and unfamiliar faces, respectively. Participants fixated to the right of a placeholder (an average face) before the target face was shown. The task was to indicate whether the face stimulus shown was familiar or not. Adapted from (van Belle et al., 2010).

2 General Methods

2.1 Overview

In this chapter I want to give a brief introduction to the main methods used in the experimental chapters. These are related to structural and functional measurements of the human brain using MRI (Magnetic Resonance Imaging). The experimental chapters will come with their own method sections and give more detail on the specific methods used for individual experiments. What follows is meant to serve as an overview and basis for these more detailed sections.

2.2 Basics of Magnetic Resonance Imaging (MRI)

2.2.1 Introduction

Magnetic resonance scanners offer the unique opportunity to obtain high-resolution images of the brain and other organs non-invasively. They even allow measuring brain activity in vivo (with certain limitations, see below). Modern machines and imaging sequences rest on discoveries leading to at least six Nobel prices in physics and medicine (McRobbie, Moore, Graves, & Prince, 2007) and are at the heart of an interdisciplinary endeavor spanning from quantum mechanics to vascular physiology. I am not an expert in any of these fields. Nevertheless, here I will try to give a basic overview of the principles underlying MRI.

2.2.2 Spins and the static magnetic field

MRI scanners mainly are big electromagnets, typically resting on super-conducting circuits and with very high field strengths. For instance, the data presented in this thesis were acquired using magnets with field strengths of 1.5 and 3 Tesla (T) (corresponding to about 3 and 6 $*10^5$ times the strength of the magnetic field of the earth). The static magnetic field of the magnet is commonly referred to as B_0 and will interact with an object placed in the magnet. Specifically, it will interact with certain atomic nuclei, in the case of biological tissue mainly with hydrogen nuclei or protons. This is because of a quantum mechanical quality of these protons called *spin*. One can think of a proton as revolving around its own axis. The spinning proton

is a charged particle and thus will have a magnetic moment along its spinning axis.

This magnetic moment in turn can have two 'states' (often referred to as 'up' and 'down'). These states can be envisioned as the sign of the magnetic moment along the spinning axis and as being determined by the spinning direction of the proton (clockwise or anticlockwise). When placed in B_0 a fraction of the (randomly oriented) spinning axes will align almost parallel to the magnetic field. The axis of B_0 is usually referred to as the *z*-axis and parallel to the bore of the magnet (the vertical and left-right axes are referred to as *y*- and *x*-axis, respectively). The spinning axes will wobble or 'precess' around the z-axis due to torque, like a gyroscope. The frequency with which spins precess around the z-axis is known as Lamor frequency and depends on the strength of the magnetic field (in the case of hydrogen nuclei 42.58 MHz/T). Finally, the spins can be aligned with or in opposition to the direction of the z-axis, depending on their state.

2.2.3 Resonance

Z-aligned or parallel orientation of the spins is a state of slightly lower energy than opposing or anti-parallel orientation. At room temperature slightly more spins will be oriented parallel than anti-parallel. This causes a net vector of magnetization that is aligned with the (much stronger) B_0 and called equilibrium magnetization or M_0. The difference in energy levels between the two states depends on field strength and is equal to the energy carried by a photon oscillating with Lamor frequency. Spins in the lower-energy state can absorb the energy carried by a wave of this

frequency and 'flip' their orientation from being aligned with to opposing B_0. This 'resonance' is what the R in MRI refers to and forms the basis of the signal measured.

Typically, Lamor frequency is in the radio frequency (RF) range. An RF pulse can be created by sending a current alternating with this frequency through a coil placed around the x-axis. This will induce a magnetic field B_1 that is parallel to the x-axis and static from the point of view of the precessing spins. Consequently the net magnetization vector will tilt towards the x-y or *transverse* plane. The resulting tilted magnetization vector is called M_1 and its angle will depend on the duration of the RF pulse. For instance, a 90 degrees RF pulse will tilt M_1 in the transverse plane, where it will rotate around the z-axis with Lamor frequency. This will result in a measurable signal of the same frequency in the RF coil (or in a separate receiver coil).

2.2.4 Transverse relaxation

The M_1 signal will quickly decay because the magnetic moments of individual spins interact. These interactions result in local inhomogeneities of the magnetic field and cause the vectors of the spins to *dephase*. That is, the spins still precess around the z-axis and in the transversal plane but fan out of phase; they do not point in a single direction at any given point in time any longer. The resulting decay of the signal is exponential and in brain tissue typically drops to 37% of the original signal intensity in around 100 ms (the exact timing will vary between tissue types). This type of signal decay is referred to as *spin-spin relaxation* or *transverse* relaxation and the

time constant describing its steepness is called T_2. In practice spins dephase faster than indicated by T_2 because spin-spin interactions are not the only source of magnetic field inhomogeneities. The effective decay time due to dephasing is thus sometimes referred to as T_2^* (the difference between T2 and T2* varies with magnet quality).

2.2.5 Longitudinal relaxation

A second, slower process of relaxation is *spin-lattice* or *longitudinal relaxation*. It refers to a process in which spins that flipped into the high-energy state pass the energy they absorbed from the RF pulse back to the surrounding tissue in the form of heat. As a result the spins flip back to the low-energy state and the net magnetization vector grows back along the z-axis (M_z). The time it takes M_z to grow to 63% of the initial equilibrium state M_0 is called T_1 and depends on field strength and the surrounding tissue. In brain tissue T_1 typically is an order of magnitude bigger than T_2 (around 1000 ms).

2.2.6 Contrast

Both, T1 and T2 vary slightly between tissue types, for instance between bone, grey matter, white matter and cerebrospinal fluid. Thus when the resonance signal is sampled at a point in time before full relaxation its intensity will depend on the type of tissue excited by the RF pulse. This is the basis of contrast in MR images. The time between

excitation and readout of the resonance signal will determine whether the contrast is weighted by differences in T1, T2 or overall proton-density.

2.2.7 Spin echo and gradient echo

In practice, the resonance signal isn't sampled directly, but rather a so-called *echo* of this signal. For instance, an initial 90 degrees RF pulse can be followed up by a second 180 degree pulse after the decay of the signal due to dephasing. This will reverse the effect of dephasing induced by field inhomogeneities in B_0, but not the dephasing induced by spin-spin relaxation or T_2. This is because the dephasing induced by magnetic field inhomogeneities is systematic– spins that were precessing faster due to this dephasing will lag behind after the second flip and vice versa. Eventually (after twice the lag between the first and second pulse) the spin vectors will be coherent again and emit an RF echo, which in this case is referred to as *spin echo*. The dephasing induced by spin-spin interactions on the other hand is random and cannot be reversed; it will go on throughout the sequence. Thus, the magnitude of the spin-echo will depend on T_2 alone and not be affected by local field inhomogeneities as in T_2*. The time between the initial RF pulse and the echo is called *TE*. The contrast of the image will depend on the difference in T_2-related decay between tissues at TE.

A second form of echo is *gradient echo*. Gradient echo sequences typically use an RF pulse < 90 degrees and an additional gradient magnetic field that will deliberately introduce dephasing of the spins. This gradient is followed by a reversed gradient causing re-phasing of the spins and consequently an echo of the signal. The second gradient only counteracts the

dephasing induced by the first gradient. Therefore the magnitude of the echo will diminish over time due to both, local field inhomogeneities and spin-spin relaxation. Thus, gradient echo images depend on T_2* contrast at TE.

2.2.8 Spatial encoding using gradients

Gradient magnets cause the overall magnetic field strength to vary along the axis on which they are used. This principle is also used to modulate the field strength for each point in space (*voxel*) and thus encode spatial position in the signal. Gradient magnets are used for all three axes and can typically be switched on and off very fast. A common way of spatial encoding involves the use of one gradient for *slice selection*. This gradient will result in a varying Lamor frequency along its axis and is left on during the excitation RF pulse. This will ensure that the Lamor frequency of the spins matches the bandwidth of the RF pulse only for a certain region or slice along the gradient. Only spins of this slice will be able to absorb the energy of the pulse and thus be excited.

To encode the in-plane locations along the other two axes two more gradients are used –a phase-encoded (*PE*) and a frequency-encoded (*FE*) gradient. The phase-encoded gradient typically is switched on and off before the read-out of the signal. This will 'speed' up or 'slow' down the spins' precession for the duration of the PE gradient pulse in a position dependent manner. After the pulse all spins will return to Lamor frequency but their phase will be position dependent along the axis of the PE. The position shift corresponding to a phase shift of 360 degrees and thus to

coherent spins will depend on the steepness of the PE. This way the PE renders the signal selective for locations corresponding to a specific spatial frequency along its axis. Typically for each excitation pulse one spatial frequency along the PE axis is probed (which corresponds to one line in so-called *k-space*).

The frequency-encoded gradient (*FE*) can be thought of as a time varying form of PE along the remaining axis. The amount of phase shift per position shift will increase over time and the signal will be sampled multiple times after the excitation. Thus all spatial frequencies along the FE axis are sampled in real time after a single excitation (corresponding to column entries of a line in k-space). The re-phasing induced by the second half of the bipolar FE gradient causes the echo that is being read-out. Thus the FE gradient is also referred to as the *readout gradient*. Finally, a two-dimensional Fourier transformation is used to transform the resulting image in k-space (corresponding to spatial frequencies along two axes) to image space (corresponding to spatial positions along these axes). The details of RF pulses and gradients will depend on the particular sequence used, but the example given should demonstrate the basic idea.

2.2.9 Echo planar imaging (EPI)

In this thesis all anatomical images were obtained using T_1-weighted sequences and functional images were obtained using T_2*-weighted gradient-echo *EPI*-sequences (Echo Planar Imaging). EPI-sequences are well suited for functional measurements because they allow very fast acquisition times and the T_2* contrast is quite sensitive to the BOLD-signal (Blood

Oxygen Level Dependent, see below). This comes at the price of high proneness to susceptibility and other artifacts. Typically, EPI sequences use a single RF pulse (single-shot) to acquire data from a whole slice.

In one version of EPI the phase-encoded gradient is persistently on during the acquisition of the whole slice. This causes a continuous drift of the signal along spatial frequencies of the PE-axis (i.e. along the lines of k-space). At the same time there is a train of bipolar FE or readout gradients with each *lobe* (reversal) of the readout gradient corresponding to one signal echo and encoding all spatial frequencies along the FE axis (i.e. a whole line of k-space). However, the continuous PE gradient causes the signal to shift along the PE axis even during the acquisition of a line in k-space, which causes a 'zig-zaggy' acquisition course in k-space. This can be avoided using *blipped* EPI. In blipped EPI the continuous PE gradient is replaced with a train of short PE 'blips' coinciding with the reversal of the readout gradient and accumulating phase shifts along the PE axis. (Hornak, 1996; McRobbie et al., 2007; Oppelt, 1983)

2.3 Voxel Based Morphometry (VBM)

2.3.1 Overview

Voxel Based Morphometry (VBM) is a method designed to quantify volumetric differences in local brain tissue between participants. The first step of VBM is to segment the image according to tissue types. These tissue type specific images are then warped to a group average and finally to an anatomical template. The process of warping will require digital 'squeezing'

or ‘stretching’ of tissue, depending on whether an area in the individual brain is bigger or smaller than the respective area in the anatomical template. The amount of ‘stretching’ and ‘squeezing’ done during normalization will be reflected in the value assigned to each voxel in VBM. Thus the resulting image for each participant is a map reflecting local gray matter volume (Ashburner, 2009).

2.3.2 Unified segmentation

The segmentation algorithm in Statistical Parametric Mapping (SPM 5 and upwards, http://www.fil.ion.ucl.ac.uk/spm/) models the intensity distributions of different tissue types as a mixture of Gaussians (Ashburner & Friston, 2005). The algorithm will estimate the mean, variance and mixing proportions of the tissue class-specific intensity distributions, which will allow to explicitly model partial volume effects (i.e. a mixture of tissues within a voxel). Based on the estimated intensity distributions, voxel intensities in the image can be interpreted as probabilistic evidence for this voxel belonging to a certain tissue class (or, in Bayesian terms, the ‘likelihood’). This evidence is then combined with priors derived from tissue probability maps to assign a voxel to a tissue class following Bayes’ algorithm.

The segmentation algorithm further estimates and removes image inhomogeneities with a discrete cosine transformation (DCT) of the image using low spatial frequency three-dimensional basis functions. Furthermore, the algorithm needs to bring the images in registration with the tissue probability (prior) maps. This is done by warping the image based on an

estimation of about 10^3 transformation coefficients for a set of cosine basis functions. All three processes – inhomogeneity correction, tissue classification and nonlinear warping of the image for registration are realized in a single generic model. This avoids arbitrary decisions on the order of these steps, that would potentially bias the results (Ashburner & Friston, 2005).

2.3.3 DARTEL

After the images are segmented the resulting grey and white matter images are registered to each other using a more elaborate registration algorithm, resting on about six million transformation parameters (*DARTEL*, 'Diffeomorphic Anatomical Registration using Exponentiated Lie algebra'; (Ashburner, 2007)). Initially the single subject images are roughly aligned with each other using a rigid body registration optimizing six parameters. DARTEL then alternates between creating an average template of all maps and warping the first level images (i.e. individual subjects) to this average.

The optimization function for the warping process aims at the minimization of differences between the images and thus contains a term that reflects image similarity. At the same time the algorithm aims at satisfying a set of priors that effectively penalize deformations that are too extreme. Large deformations can break the required one-to-one symmetry of points between the warped and target images. DARTEL avoids this by decomposing large deformations into a sequence of smaller deformations each of which satisfies one-to-one symmetry. The resulting map of

deformations can be represented as a 'flow field' that indicates the necessary image transformations.

The final average of images can be registered to an anatomical template, which allows anatomic labeling in a standard stereotactic space. Also, the images can be *modulated* with the so-called Jacobinian determinants. These numbers correspond to the relative volumes of tissue before and after warping. So when a region was warped to double its size to fit the template the corresponding intensity values in the registered image would be divided by two. Thus the final intensity values in the registered image are a proxy to the grey matter volume of the original image in this area.

2.3.4 Analysis

The registered maps containing grey matter volume values can be spatially smoothed (to remove high spatial frequency noise) and analyzed statistically. For instance, one can define a region of interest in template space and retrieve the sum of voxel intensities from this area. This is done for each registered gray matter image and thus yields an index of local gray matter volume for each participant. These individual gray matter volumes can then be compared between groups or correlated with other participant-specific traits. (Ashburner, 2009)

2.4 The Blood Oxygen Level Dependent signal (BOLD)

2.4.1 Physics of BOLD

Functional MRI (fMRI) typically measures the so-called *BOLD* (Blood Oxygen Level Dependent) signal. BOLD rests on a difference between oxygenated and deoxygenated hemoglobin with regard to their magnetic properties. The iron atom of hemoglobin has unpaired electrons in the deoxygenated state of the molecule and consequently renders it paramagnetic. Thus the presence of deoxygenated hemoglobin will introduce local inhomogeneities in a magnetic field like B_0. Oxygenated hemoglobin on the other hand is diamagnetic and does not interact much with the magnetic field (in oxygenated hemoglobin the electrons of the iron atom are paired with those of the oxygen; (Pauling & Coryell, 1936)). So the magnetic properties of blood depend on its oxygenation and changes in blood oxygenation can modulate local homogeneity in a magnetic field. This in turn will affect T_2* weighted MR signals (see above).

2.4.2 Vascular physiology

The BOLD signal is a vascular, rather than a neural signal. It depends on the paramagnetic properties of deoxygenated hemoglobin and is thought to arise from changes in blood oxygenation in capillaries and venules. In practice, larger vessels contribute more to the measured signal at low field strengths and for gradient- vs. spin-echo sequences (Logothetis, 2008).

Given deoxygenated hemoglobin is paramagnetic and causes faster T_2* it is somewhat surprising that neural stimulation is usually followed by a

positive BOLD signal. Neural stimulation in this case refers to e.g. the presentation of visual stimuli or direct stimulation of excitatory neurons (Lee et al., 2010). The basis of this is thought to be an overcompensating supply of oxygenated blood following neural activity.

A popular model of this process (Buxton, Uludağ, Dubowitz, & Liu, 2004) proposes that neural activation is followed by a local increase in cerebral blood flow (CBF), that surpasses a parallel increase in the local cerebral oxygen consumption rate (CMRO2). Note that the rate of oxygen uptake has an upper limit because it rests on diffusion and that the increased CBF will also serve to increase the supply of other substances like glucose. Increased CBF and CMRO2 in turn result in a *de*crease of the oxygen extraction rate and thus the local concentration of deoxygenated hemoglobin. At the same time local cerebral blood volume (CBV) increases. Finally, it is the combination of increased CBV and decreased concentration of deoxygenated hemoglobin that yield the measured BOLD response.

Typically the BOLD response to an impulse stimulus is delayed by 1-2 s relative to stimulus onset and has a width of 4-6 s after which it is followed by an undershoot below baseline that can last up to 30 s. Sustained stimulation goes along with a ceiling response, and the response to multiple stimuli is typically captured by linear superposition as long as the stimulus duration is not too short (<<4 s). Sometimes the positive BOLD response is preceded by an *initial dip*, reflecting an initial increase in deoxygenated hemoglobin. This might be due to a steeper initial rise of CMRO2 than CBF.

The post-peak undershoot in turn could reflect a 'balloon' like effect of vessels. According to this idea CBV returns to baseline more slowly than

CBF does. The venous outflow lags behind the previously increased influx and thus the 'inflated' venules are sluggish in their return to the initial diameter (Buxton et al., 2004).

2.4.3 Neurovascular coupling

The motivation for fMRI in most cases (and in this thesis) is to learn about neural activity and mechanisms rather than about vascular effects per se. Thus the vascular signal is used as a proxy for neural activation and an exact understanding of the link between neural activity and vascular responses would be highly valuable. Unfortunately, however, the BOLD signal is ambiguous with regard to underlying neural processes and the mechanisms of neurovascular coupling are still an active area of research (c.f. (Logothetis, 2008)). Even the link between brain metabolism and BOLD is indirect and mediated by a signaling cascade, rendering energy consumption and BOLD somewhat dissociable (Attwell & Iadecola, 2002).

A direct way to investigate neurovascular coupling is to obtain BOLD and electrophysiological data simultaneously. Several animal studies recorded BOLD signals, local field potentials (LFP) and multi-unit activity (MUA) in parallel. MUA is found by high pass filtering the signal of electrophysiological recordings (e.g. > 1 KHz) and thought to reflect the spiking of small neural populations in the vicinity of the electrode tip (up to several hundred micrometers). LFP is found in low frequency bands of the signal (e.g. <300 Hz) and thought to reflect 'perisynaptic' activity up to several mm from the electrode. This perisynaptic activity is thought to reflect all sorts of neuromodulatory and subthreshold processes including

calcium-mediated post-spike membrane oscillations (afterpotentials). LFP seems to reflect mainly modulatory input and intracortical processing rather than pyramidal output (Logothetis, 2008).

The BOLD response has been found to be more tightly correlated with LFP than MUA (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001). Specifically, most variance of BOLD responses to visual stimuli in awake monkeys is explained by LFPs in a frequency band from 20 to 60 Hz (Goense & Logothetis, 2008; Magri, Schridde, Murayama, Panzeri, & Logothetis, 2012). More importantly, MUA quickly adapts to sustained stimulus presentations while LFP and BOLD both persist (Goense & Logothetis, 2008). Therefore BOLD is interpreted to reflect LFP more directly than MUA.

(Logothetis, 2008) outlines candidate neural network activities underlying BOLD in terms of a *canonical microcircuit*. This microcircuit can be thought of as a column or slab of cortex with glutamatergic excitatory cells in superficial and deep layers and gabaergic inhibitory cells in middle layers. Two principle forms of activity in this exhibition-inhibition network are 'up' and 'down' states, which are characterized by more or less overall activity, but both come with a balance between excitatory and inhibitory cells. This is thought to go along with an increase and decrease in BOLD signal, respectively (note that there will be no stimulus associated output of the network in either case). A third possible state of the network is net excitation which results in higher output and is thought to be accompanied by an increase in BOLD signal as well (Lee et al., 2010), but c.f. (Logothetis, 2010).

The hemodynamic correlate of a fourth and final state of net inhibition is less clear. In this context the phenomenon of sustained negative BOLD responses is relevant. At least in some cases such a negative BOLD response points to an underlying reduction of LFP and MUA below baseline, e.g. (Shmuel, Augath, Oeltermann, & Logothetis, 2006); c.f. (Moraschi, DiNuzzo, & Giove, 2012). This might be due to net inhibition or overall down-regulation of the local networks. However, net inhibition has also been shown to come with increased energy consumption and thus might be accompanied by positive BOLD signals in some cases (Logothetis, 2008); but see (Attwell & Iadecola, 2002).

The difficulties regarding the interpretation of BOLD signals have been driven to a further level by a more recent study finding hemodynamic signals in the absence of *either* MUA *or* LFP (Sirotin & Das, 2009); c.f. (Ekstrom, 2010). In this study the authors measured MUA, LFP and hemodynamic signals in macaque V1, using optical imaging to determine CBV and blood oxygenation. The hemodynamic signals were closely related to stimulus evoked MUA and LFP. Critically the hemodynamic response was also seen when the monkeys were merely prompted in a periodic fashion to fixate without stimulating the surrounding visual field at all (the hemodynamic signal was obtained from recording sites outside foveal representations). This anticipatory hemodynamic signal was not accompanied by either the recorded LFP or MUA. This renders the supposed neural basis of this signal unclear. The authors speculated that it might rely on distal neruomodulation of arteries (Sirotin & Das, 2009).

In a follow-up study (Cardoso, Sirotin, Lima, Glushenkova, & Das, 2012) the authors decomposed stimulus-evoked and task-related components of hemodynamic signals in response to varying stimulus contrasts. They found that hemodynamic stimulus responses were linearly related to stimulus evoked MUA which in turn had a non-linear relationship to stimulus contrast. Crucially, the task-related component could not be explained by stimulus contrast or stimulus evoked MUA and dominated the hemodynamic response. The authors speculated that the task-related component of the hemodynamic response might rest on ascending or neuromodulatory activity or corticocortical feedback from higher visual areas. This hypothesis would fit with the standard model presented above and with a relationship between the hemodynamic signals and LFP (Friston, 2012). Unfortunately the authors do not report results from LFP frequency bands in this latter study and only note 'that the goodness of fit between sensory hemodynamics and gamma-band LFP was much more variable than with spiking' (Cardoso et al., 2012).

2.4.4 Conclusion

Taken together the BOLD signal opens a unique window to study brain activity in vivo, non-invasively and with relatively high spatial resolution. However, it is an indirect measure of neural activity and its neuronal causes are unclear. It seems dominated by presynaptic, modulatory activity and its closest electrophysiological match is LFP (rather than spiking output or MUA). In line with this, BOLD seems especially sensitive to modulatory cognitive processes like attention or stimulus

awareness, e.g. (Gandhi, Heeger, & Boynton, 1999; Haynes & Rees, 2005b); c.f. (Logothetis, 2008). However, on its own the BOLD signal is merely an indicator of *some* neural activity at a given location without providing much information about the *kind* of this activity. As (Logothetis, 2008) points out this is a common problem of aggregate measures of neural activity (including MUA and LFP) and it renders the use of complimentary methods in neuroscience mandatory.

2.5 Analysis of functional MRI data

2.5.1 Introduction

The rapid nature of the EPI sequence allows acquiring many images of the brain within a short timeframe (see above). Such a time-series of data is expected to reflect dynamic changes in the BOLD signal and the underlying neural activity. Typically a time-series is analyzed by fitting the parameters of a model to the data of each voxel independently. The researcher builds a forward model of expected activations on the basis of knowledge about the time-course of stimuli, conditions and nuisance parameters. The parameters of the fitted model are then tested for statistical significance against zero. The resulting test statistic for each voxel (e.g. *t*-values) can be represented as a new image, a so-called *statistical parametric map* (SPM). A popular software package for this type of analysis is Statistical Parametric Mapping, (SPM, http://www.fil.ion.ucl.ac.uk/spm/), which is written in the Matlab programming environment

(http://www.mathworks.co.uk/products/matlab/). SPM 8 was used for at least part of the analysis of all MRI experiments presented in this thesis.

2.5.2 Preprocessing

The raw time-series of a given voxel will contain variance that is not due to the BOLD signal. This variance can be much greater than the signal of interest and part of it can be systematic (e.g. due to head motion). It is therefore necessary to account for these sources of error before fitting a model to the data. This *preprocessing* of the data comprises several standard steps that are common to all experiments presented in this thesis (Ashby, 2011; Poldrack, Mumford, & Nichols, 2011).

First, the images are *mean-bias corrected* for gross, low spatial frequency variations. Most of the functional MRI experiments presented in this thesis were conducted with a 32-channel RF-coil stripped to only 20 channels at the back of the head to avoid restrictions of the field of view. Thus the intensity of raw images was spatially heterogeneous, rendering this step of preprocessing important. The heterogeneity of the images is estimated based on the first image of the time-series using image segmentation and a set of low spatial frequency basis functions (see above, unified segmentation). An inverted version of this heterogeneity image can then be applied to the time-series to even out the bias.

The next step of preprocessing is to correct the image for distortions due to heterogeneity of B_0. These field inhomogeneities are most prominent at borders between tissue and air and can be measured by acquiring data for a *field map*. A field map is calculated as the phase difference between two

images acquired with different echo times. These differences can be used to estimate the spatial shift of each voxel in the EPI images, *a voxel displacement map* (VDM). Finally, the distortions in the image can be corrected by applying an inverted version of the VDM to the images.

Typically, participants in an fMRI study have their head tightly padded with foam cushions to avoid head motion. Despite this restriction some residual motion will affect the data and induce variance in the time-series of a voxel. This effect can be dramatic, e.g. at the grey matter / white matter boundary where the tissue types falling within a voxel can change due to head movements. Thus the images of the time-series have to be *realigned*. SPM uses a so-called rigid-body transformation for realignment. One image of the time-series is used as a template reference and the remaining images are matched to this reference using a six-parameter transformation. Under the assumption of constant head shape, the motion-induced difference between two images can be described by a translation parameter and a separate rotation parameter for all three axes of space. The parameters are estimated using an iterative optimization algorithm aiming at minimizing the sum of squared differences (SSD) between the images. Such optimization algorithms are prone to settle for local minima, rendering it important to manually correct for gross differences between images (gross differences can arise e.g. for data from different sessions).

EPI images typically have a lower spatial resolution (e.g. 2.3mm isotropic voxels) than structural, T1-weighted scans (e.g. 1mm isotropic voxels). Thus anatomical structures can be identified with greater precision for the structural scan and it is desirable to *co-register* the functional data to

this high-resolution image. The optimization criterion for co-registration is less straightforward than for realignment. SSD would not work in this case because functional and structural images differ in the (sequence-dependent) image intensities evoked by different tissue types. Therefore co-registration in SPM aims at maximizing *mutual information* or reducing the *joint entropy* of images. Voxels are binned with regard to their intensity values in both images, resulting in a two-dimensional joint histogram. The algorithm will then search for parameters of a rigid-body transformation that yield a joint histogram indicating minimal joint entropy. That is, it will search for an image transformation that yields a configuration in which knowledge about the intensity bin of a given voxel in one image will reduce uncertainty about its intensity bin in the other image maximally. This will typically be the case when the images are well aligned.

Finally, the data are spatially *smoothed*. EPI images will contain random noise with high spatial frequency and the BOLD signal is expected to be somewhat spatially coarse. Thus a weighted average of the local signal will yield a better signal-to-noise ratio than the raw data. Smoothing in SPM is done using a three-dimensional spatial Gaussian filter.

These are the main preprocessing steps employed for the MRI experiments presented in experimental chapters. There are, however, some additional steps not mentioned yet. *Temporal high-pass filtering* is used to strip the data of very slow fluctuations that are most likely caused by temperature dependent scanner drifts. In SPM this is implemented in the model-estimation step: A set of discrete cosine basis functions capturing the low frequency components of the signal are entered into the design matrix

as regressors of no interest (see below). *Segmentation* and *normalization* are optional steps of preprocessing that have already been discussed in the context of VBM (see above).

2.5.3 General linear model

2.5.3.1 Overview

Probably the most popular model in fMRI analysis is the *general linear model* (GLM; (Ashby, 2011; Poldrack et al., 2011)). The general form of which is

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

Where *Y* is a data vector, e.g. the time-series of a voxel with *i* entries corresponding to the number of images acquired; *X* is a *design matrix* with i rows and *j* columns, the latter corresponding to the number of *regressors*, or factors used to explain the data; β is a vector of weights assigned to the j regressors and ε is a vector of i prediction errors. The design Matrix X represents a forward model of the expected BOLD signal based on the stimulus time-course and other factors. Typically, this is specified in a two-step procedure. The researcher first defines a model of neural activation and then convolves this with a hemodynamic response function (HRF, see above). For instance, a regressor for neural activation provoked by a given stimulus type might model this activation in a boxcar fashion with the neural response set to 1 whenever the stimulus is present and 0 when it is not. This series of 'boxcars' is then convolved with the assumed HRF. Note

that the design matrix is defined at a higher temporal resolution then TR and then down-sampled.

2.5.3.2 Beta estimates and underlying assumptions

Modelling the data as a linear combination of regressors rests on the assumption that the BOLD signal can be treated as stemming from a linear, time-invariant system (LTI). That is, that the BOLD signal evoked by a train of stimuli, by stimuli of different durations and by combinations of simultaneous stimuli can be modeled as a superposition of the response evoked by a single impulse stimulus. This seems to be reasonably accurate as long as stimulus durations and inter-stimulus intervals are no shorter than about 1-2 seconds (Ashby, 2011; Boynton, Engel, Glover, & Heeger, 1996; Friston, Josephs, Rees, & Turner, 1998; Pfeuffer, McCullough, Van de Moortele, Ugurbil, & Hu, 2003; Vazquez & Noll, 1998).

As X and Y are known, β can be estimated in a way that minimizes ε, e.g. via the ordinary least squares method (OLS). The β weights of individual regressors and regressor contrasts can then be tested for statistical significance against 0 (see below). Note that the fitting and statistical testing of β weights is done for the time-series of every voxel independently. Also note that β can be determined uniquely[2] and only represents an *estimate* with regard to generalization of the model beyond the sample data.

[2] At least as long as there are more independent observations than regressors and the regressors are linearly independent

Such estimates of β rest on several assumptions regarding ε. Namely, that the errors will be randomly distributed about 0, that their variance will be equal across values in X and that they will not be correlated across scans. The latter requires that the model captures all systematic changes in the data. This is almost never the case; knowing the entry of ε for one TR typically reduces uncertainty about ε for other images. Temporally high-pass filtering the data counteracts some of this (see above). SPM tries to take the remaining *autocorrelation* of errors into account with a three-step procedure. First, the errors are estimated based on an initial solution for β. Then a first-order autocorrelation model (AR(1)) is applied to the estimated errors. Finally, the resulting covariance matrix of the error is inverted and multiplied with the design matrix and the data to *pre-whiten* them, effectively removing (most of) the unexplained autocorrelation. The final estimate of β can then be calculated based on the pre-whitened version of data and design matrix.

As mentioned above, the design matrix can contain regressors that are not stimulus related, e.g. a constant modelling baseline activation, estimated motion regressors or slow frequency phasic functions to capture drift in the data. The weights assigned to the individual regressors (the entries of the β vector) will reflect the variance in the data that is uniquely captured by the respective regressors. So entering nuisance regressors in the model will prevent any variance that can be explained by these from being assigned to regressors of interest.

2.5.3.3 Statistical inference

Under the assumptions of the GLM regarding the error (see above) the standard error of each beta estimate or contrast of beta estimates can be estimated as well. This in turn allows computing a t-statistic for each voxel as

$$t = c'\beta/\mathrm{SEM}(c'\beta)$$

where *c* stands for a vector specifying the respective contrast of beta weights (e.g. [1 0 0…] would be used to test the first entry of β for significance), SEM indicates the corresponding standard error and *t* stems from a t-distribution with degrees of freedom equaling the number of images minus the number of parameters in β.

The corresponding SPM will contain as many *t*-values as there are relevant voxels in the image and consequently the threshold for determining statistical significance needs to be adjusted for multiple testing. A simple Bonferroni correction would come with a loss of sensitivity and is therefore impractical. SPM circumvents this problem by taking into account the spatial smoothness of the data. The spatial correlation of image intensities renders the number of independent observations less than the number of relevant voxels. *Random field theory* (RFT) allows estimating the number of independent volume elements or *resels* in the image and computing the probability of any smooth peaks exceeding a given threshold under the null hypothesis (Worsley, 2007). This threshold can then be adjusted according to the desired family-wise error rate (FWE).

An extension of RFT based statistics is *cluster-level inference*. Cluster-level inference requires choosing an arbitrary cluster-forming threshold and then estimating the probability of a contiguous cluster of activation above this threshold occurring under the null hypothesis. Crucially, this probability can be expressed as a function of cluster size. Thus one can set the cluster level FWE by only considering clusters of activity as statistically significant that exceed both the cluster forming threshold and the corresponding minimum cluster size.

If the data are spatially normalized the researcher can compute SPMs for the whole sample (i.e. at the *second level*) instead of doing this at the level of individual participants (i.e. at the *first level*). First-level SPMs only take into account the intra-individual error variance (the *fixed effects* variance) and thus inference is restricted to the specific participant. Second-level SPMs on the other hand allow taking into account the variance between participants, referred to as *random effects*. Typically, second-level analyses are *mixed effects analyses* that take into account both aspects of variance and thus allow inferences about the population from which participants are drawn. This can be done by iteratively estimating the within-subject variance and mean effect for each participant as well as the between subject mean effect size and variance. SPM greatly simplifies this problem by doing a second-level *t*-test on the first-level contrast images. Note that the empirical variance between first-level contrasts will reflect both, within and between subject variance. Assuming equal intra-individual variance across participants (or at least ignoring heterogeneity) allows to

simply use the empirical variance as an estimate for the total mixed effects variance.

The analyses described so far are *mass-univariate* and typically aim at signal detection or at comparison of amplitudes. However, other aspects of the data might be of interest. Among them are the separability of patterns of activation evoked by different stimuli and the feature tuning of a voxel's response. In this thesis I used two further analysis techniques to investigate such aspects of the data that will be introduced below: *Multivoxel Pattern Analysis* (MVPA) and *retinotopic mapping*.

2.5.4 MVPA

2.5.4.1 Overview

Multivoxel Pattern Analysis (MVPA) poses the question whether different types of stimuli evoke distinct *patterns* of brain activity, e.g. (Haynes & Rees, 2006; Norman, Polyn, Detre, & Haxby, 2006). Typically such analyses are confined to a region of interest for which some level of BOLD activation is expected. For instance, MVPA can be used to distinguish patterns of activity in early visual cortex evoked by different stimuli – and their modulation via attention or stimulus awareness (Haynes & Rees, 2005a; Kamitani & Tong, 2005). Thus MVPA is not limited to the detection of changes in overall signal amplitude for a given area but allows asking about the stimulus related information carried by patterns of activation.

One way of conducting MVPA is to obtain a set of SPMs for each stimulus category of interest and to use these as pattern exemplars.

Typically, *t*-maps provide a more stable estimate of the evoked pattern than beta or contrast maps because they take the error into account (Misaki, Kim, Bandettini, & Kriegeskorte, 2010). Such *t*-maps can then be masked by a region of interest and flattened into a single column of data with each entry corresponding to a voxel in the original image. Geometrically these columns can be thought of as *vectors* in an *n-dimensional feature space* spanned by the n voxels of the region of interest. The question for stimulus related information in the patterns then becomes a question of spatial separability of these vectors.

2.5.4.2 Classification algorithms

Vectors corresponding to a stimulus category will typically show some within-category scatter. However, the multivariate distance between category means might still allow distinguishing vectors by category above chance. C*lassification algorithms* are tools for testing this. They aim at optimizing a multivariate decision boundary to distinguish categories in feature space. For instance, *Linear Discriminant Analysis* (LDA) explicitly aims at maximizing the ratio of within vs. between category variance. Under the assumption that the data from two categories stem from multivariate normal distributions the decision boundary or hyperplane satisfying this criterion will have a normal that is the vector connecting the category centroids. This can easily be imagined for the two-dimensional case: The line that will best divide two normally distributed groups of data points in a plane will be orthogonal to the line connecting their multivariate means. LDA is a computationally efficient way of classification but cannot cope with

datasets containing fewer observations than dimensions. Given the typically high number of voxels in regions of interest fMRI data requires some form of *dimensionality reduction* before applying LDA. This can be achieved e.g. via a principal component analysis (PCA). Support vector machines (SVMs) are computationally more demanding than LDA but can cope well with high-dimensional datasets. SVMs place the decision boundary in a way that will maximize its margin, i.e. its distance to the closest vectors on either side of the boundary (the support vectors). Linear SVMs also tend to generalize well across datasets, which is an important sign of validity for any classification algorithm (see below).

2.5.4.3 Generalization

In principle almost any two groups of vectors in feature space are separable. A sufficiently complex decision boundary should be able to separate data that is randomly assigned to an arbitrary number of categories. This highlights an important distinction: Classification algorithms address questions about *in-principle* separability of categories that generalize beyond the specific dataset used to train the algorithm. Therefore algorithms have to strike a balance between minimizing the classification error with regard to the dataset at hand and finding a parsimonious decision boundary that will not *overfit* the peculiarities of this dataset. With regard to fMRI linear SVMs and LDA algorithms tend to do this reasonably well; c.f. (Misaki et al., 2010). However, the generalizability of the classification needs to be tested explicitly. A common way of doing this is the *leave-one-out approach* in experimental design and analysis.

The leave-one-out approach entails scanning a participant for multiple runs containing at least one trial for each stimulus category to be classified. This allows setting up separate GLMs for each run, which can then be used to obtain one *t*-map (and thus vector) per stimulus category and run. The data is then split into a *training dataset* containing the data of all but one run and a *test dataset* comprising of the data from the remaining run. The training dataset is fed to the classification algorithm, which aims at optimizing the decision boundary. Finally, the test dataset is used to probe the classification performance this boundary achieves and the corresponding *classification accuracy* is stored. This procedure is iteratively repeated until each run served as the test dataset once.

The resulting proportion of correct classifications can be tested for statistical significance vs. chance level using a binomial test. Across participants the sample distribution of average classification accuracies can be tested vs. chance level using a *t*-test. Classification accuracies for different conditions can be compared using e.g. analysis of variance (ANOVA) or pair-wise *t*-tests. Comparisons across participants or regions of interests are also possible but more problematic because of possible confounds. For instance, classification accuracy might increase for bigger regions of interest without necessarily indicating any difference beyond this confound.

2.5.4.4 Searchlight and representational similarity analyses

MVPA can also be used to *detect* regions of activation carrying stimulus information (Kriegeskorte, Goebel, & Bandettini, 2006). This is done using a so-called *searchlight* approach. The algorithm first defines a

seed voxel and then uses MVPA to determine classification accuracy based on patterns of activation surrounding this voxel. For instance, it might consider patters of activation in all grey matter voxels that fall within a sphere of fixed radius centered on the seed voxel. The resulting classification accuracy (minus chance level) is then assigned to the seed voxel and the whole procedure repeated for every voxel of grey matter. The result is an *accuracy map*, which can then be normalized and subjected to second level statistical analyses just like the contrast images of a GLM (although cluster level FWE correction is overly conservative in this context, (Stelzer, Chen, & Turner, 2013)).

Apart from classification, multivariate analyses can be used to e.g. test the reliability of evoked patterns across trials or investigate the similarity of patterns evoked by different stimuli (Representational Similarity Analysis, RSA; (Kriegeskorte & Kievit, 2013; Kriegeskorte, 2009)). The latter allows for comparisons of neural representations across participants, measuring modalities and even species without the need for further normalization. The (dis)similarity of patterns of activity for different stimuli can e.g. be represented in a correlation matrix. It is thus transformed into an abstract space that is independent of measuring modality and neural structure.

2.5.5 Retinotopic mapping

2.5.5.1 Visual field maps

Two fundamental properties of visual cortex neurons are spatial tuning and their arrangement in *visual field maps* (Wandell, Dumoulin, & Brewer, 2007). Typically, the response of a visual neuron is restricted to stimulation of a small part of the visual field referred to as the neuron's *receptive field*. Responses tend to be maximal for stimuli at the center of the receptive field and gradually fall of with spatial distance, often below baseline for stimuli falling within in a suppressive surround (Cavanaugh, Bair, & Movshon, 2002; Webb, Dhruv, Solomon, Tailby, & Lennie, 2005). The location and size of receptive fields gradually change along the cortical surface with neighboring neurons showing similar spatial tuning and the overall arrangement forming a map of the visual field (see below).

The gradual nature of this change renders it apparent on a macroscopic scale. In humans, visual cortex spatial tuning has first been described in a systematic fashion for lesioned soldiers after the Russo-Japanese war in 1904/1905 and World War I (Inouye, 1909 as cited in (Glickstein & Whitteridge, 1987); (Holmes, 1918)). The location of occipital lesions systematically mapped onto the visual field location of scotomas. Damaged tissue at the occipital pole yielded foveal scotomas while lesions further along the calcarine sulcus resulted in a more peripheral loss of vision. Similarly, the polar angle of scotomas mapped onto the cortical surface. Lesions of the left hemisphere led to scotomas in the right visual hemifield and vice versa. Scotomas in the upper and lower visual hemifield

were brought about by lesions of the inferior and superior banks of the calcarine sulcus, respectively.

Later, MRI based lesion mapping was used to refine these early descriptions of the visual field map in human V1. Specifically, this data demonstrated that the *cortical magnification factor* (CMF) had previously been underestimated (Horton & Hoyt, 1991). Representations of the fovea occupy far more area of cortical surface per degree visual angle than more peripheral representations do.

2.5.5.2 Phase encoded retinotopic mapping

The biggest change in mapping was brought about a few years later with the introduction of retinotopic mapping using functional MRI and digital reconstruction of the cortical surface. This allowed mapping the spatial preferences of visual cortex in healthy humans with high spatial resolution. It provided an accurate description of V1 properties on the individual level and led to the discovery of numerous visual field maps in extrastriate cortex (DeYoe et al., 1996; Engel et al., 1994; Sereno et al., 1995; Wandell et al., 2007).

A typical retinotopic mapping experiment requires participants to fixate the center of a screen while dynamic high contrast stimuli travel the visual field in a phasic manner. Most commonly, an expanding or shrinking ring carrier is used to vary stimulus eccentricity and a turning wedge carrier to vary polar angle (**Figure 2-1 A & B**). Cycling through such a paradigm multiple times during an fMRI session yields a specific prediction for the time-course of the BOLD response.

Any voxels with a spatial preference that falls within the stimulated area will respond in a phasic manner.[3] Specifically, the time course of such voxels will show peaks corresponding to the cycling frequency of the stimulus. Most importantly the phase lag of these responses will reflect the spatial preference of the respective voxels. The phasic nature of the stimuli ensures that different regions of the visual field are stimulated with similar frequency but different eccentricity and polar angle bands will be stimulated at different phases (**Figure 2-1 C**). The elicited response will only be offset by a constant hemodynamic lag that can be accounted for by e.g. averaging the results of sessions with opposite cycle directions (counter vs. clockwise wedge rotation and expanding vs. shrinking ring). Applying a fast fourier transformation (FFT) to the time course of each voxel allows to determine responsive voxels as those with increased relative power at the stimulus frequency and to determine the phase lag of their response at this critical frequency.

In order to visualize the results of this analysis the inferred preference for polar angle and eccentricity can be color-coded and assigned to the respective voxels. This effectively produces a map for either parameter. However, simply color-coding the voxels would not yield a very useful map. The gradual change of spatial preference is along the cortical *surface*, which is obscured in volume images due to cortical folding. This problem is solved by digitally reconstructing the cortical surface via tessellation of the grey-matter / white-matter boundary that is determined

[3] That is any voxels containing a population of visual neuron's with a corresponding average spatial preference

in the segmentation process (see above). This allows for three-dimensional rendering of the image and digital manipulations such as flattening or inflating the cortical surface (in this thesis I used FreeSurfer for surface based analysis, http://surfer.nmr.mgh.harvard.edu).

The spatial preference of voxels can then be color-coded and projected onto the unfolded cortical surface. This yields a clearly visible parameter map and enables delineating individual visual field maps on the cortical surface. For instance, in a given hemisphere V1 can be identified as a map of the contralateral hemifield stretching from the upper vertical meridian of the visual field represented on the lower bank of the calcarine sulcus to the lower vertical meridian on the upper bank (**Figure 2-1 D-F**). V1 is flanked by V2 and V3, which are split into two quarter fields each and the boundaries between these areas are marked by a mirror reversal of changes in polar angle preference (Wandell et al., 2007).

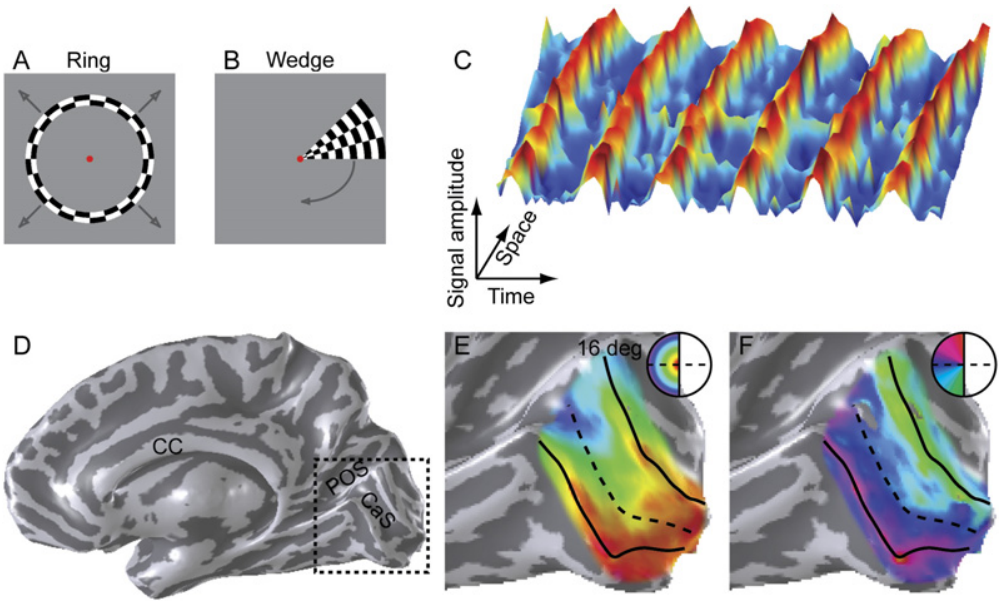

Figure 2-1 Phase-encoded retinotopic mapping. **A** Ring carrier used to stimulate different eccentricity bands in a systematic fashion. Arrows indicate that the ring is expanding, which usually is done in a step-wise manner with steps occurring at a rate divisible by TR. **B** Wedge carrier used to stimulate different polar angles in a systematic fashion. The arrow in the example indicates clockwise steps of rotation. **C** Example of a travelling wave of BOLD activation, which could correspond to the stimulation sequence shown in **A** or **B**. Signal amplitude is shown as a function of time and position on the cortical surface (labeled 'Space'). The amplitude is clearly modulated with a frequency corresponding to the stimulation frequency. The phase of the response is shifted along the cortical surface. **D** Inflated reconstruction of the cortical surface of a right hemisphere; CaS: calcarine sulcus; POS: parietal-occipital sulcus; CC: corpus callosum. The rectangle of dashes indicates the area shown in **E** and **F**. **E** Eccentricity map on the inflated calcarine. Colors correspond to eccentricities up to 16 degrees as indicated in the inset. **F** Polar angle map on the inflated calcarine. Colors correspond to polar angles of the left visual hemifield as indicated in the inset. From (Wandell et al., 2007)

2.5.5.3 Population receptive fields

A relatively recent development in retinotopic mapping is population receptive field (pRF) mapping (Dumoulin & Wandell, 2008). This method rests on forward modelling a voxel's receptive field and the resulting time-

course of its response. The pRF can e.g. be modeled as a symmetrical two-dimensional Gaussian hull. The procedure then starts with a guess concerning the basic parameters of the model, in this case the center position and standard deviation of the Gaussian. Based on these parameters and the known time-course of the stimulus the model gives a prediction regarding the time-course of neural activation. This prediction in turn is convolved with a hemodynamic response function, yielding a prediction regarding the time-course of the BOLD response. Finally, this prediction is compared to the empirical time-course and an iterative algorithm is used to fit the parameters of the model, minimizing prediction error (**Figure 2-2**).

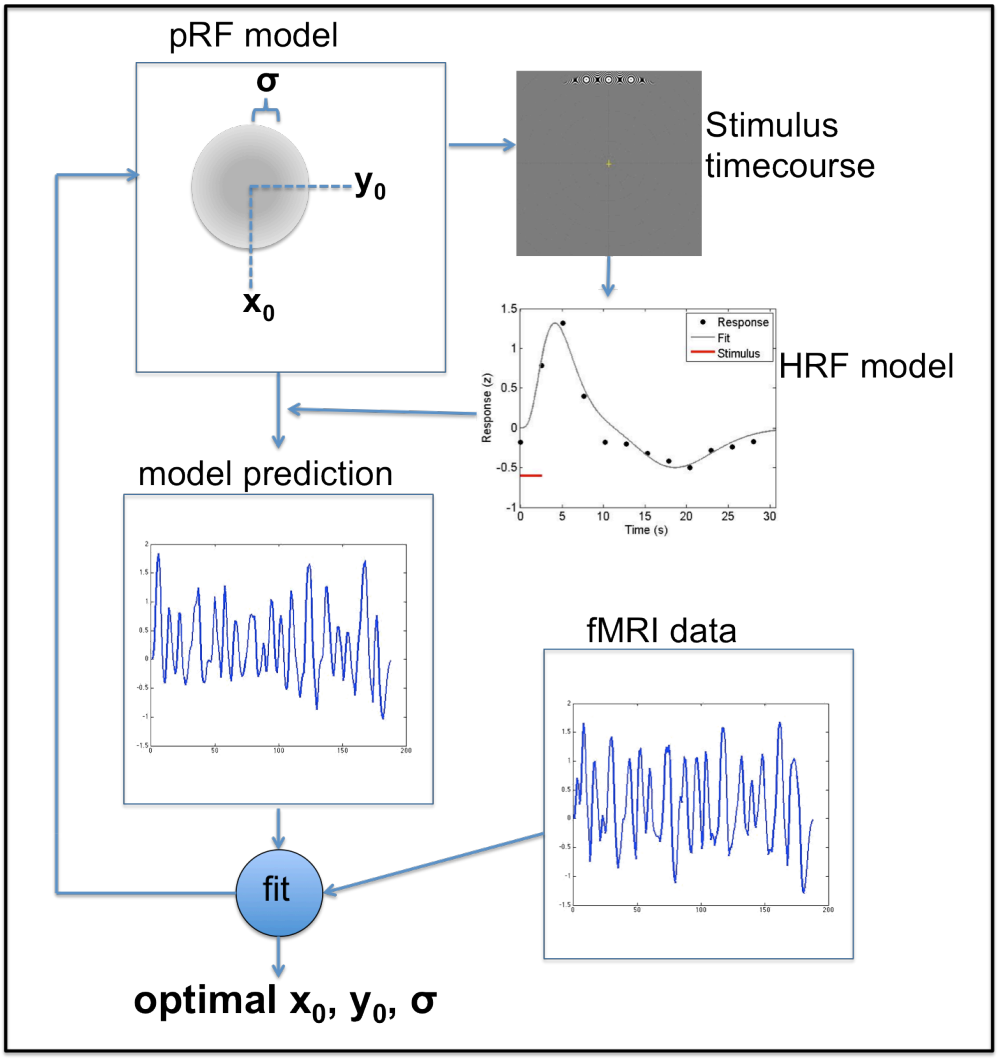

Figure 2-2 Population receptive field mapping. A hypothetical pRF with two parameters for its center position (x_0, y_0) and one for size (σ) is shown in the top left box. Based on the model and the stimulus time-course the overlap between the pRF and stimulus can be calculated as well as the predicted time-course of neural activation (top right). This prediction is then convolved with a hemodynamic response function (that has been acquired empirically in the example; middle right). The resulting model prediction (middle left) is compared to the empirical time-course of BOLD activation (lower right). The resulting prediction error is used as criterion by an iterative optimization algorithm fitting the parameters of the model (lower left). Adapted from (Dumoulin & Wandell, 2008).

Crucially, this approach will result in three parameter maps.[4] Phase encoded analyses only determine the location of maximum spatial sensitivity for a voxel, typically encoded as eccentricity and polar angle maps. pRF mapping on the other hand aims at characterizing the voxel's receptive field including its width (σ) which can be visualized across voxels as a third map. The estimated size of pRFs increases with eccentricity and along the visual hierarchy in a manner that is consistent with results from electrophysiological studies (Dumoulin & Wandell, 2008).

In principle, pRF mapping is not tied to a specific underlying receptive field model and has been successfully applied using e.g. a difference of Gaussians (DoG) model (Zuiderbaan, Harvey, & Dumoulin, 2012). An entirely different approach to pRF mapping is to first maps out the spatial preference of a voxel in a model free fashion using reverse correlation techniques. In a second step the parameters of a pRF model can be fitted to such a map for each voxel (Lee, Papanikolaou, Logothetis, Smirnakis, & Keliris, 2013).

A recent study demonstrated that BOLD-signal based estimates of pRFs in V1-3 are in good agreement with pRF properties found using electrocorticography in humans (ECoG; (Winawer et al., 2013)). The ECoG response to flickering stimuli showed a time-locked component (i.e. of a frequency matching the stimulus flicker rate) and a more broadband

[4] For a simple two-dimensional Gaussian model of the pRF. See below for alternatives.

component reflecting increased general response power during periods of stimulation. Estimates of pRFs based on either response component were highly similar with regard to center position and size. However, the amplitude of these response types differed with regard to their spatial summation properties. Increasing stimulus width led to an additive increase of response amplitude for stimulus locked responses while this increase was subadditive for the broadband component. Interestingly, pRF estimates based on BOLD-signal data show subadditive spatial summation as well (c.f. (Kay, Winawer, Mezer, & Wandell, 2013)). The authors conclude that the BOLD-signal reflects the neural response feeding into the broadband component of the ECoG signal, rather than the immediate stimulus locked response (Winawer et al., 2013).

3 Can the duration of a visual percept be modulated by the duration of a co-occurring sound?

3.1 Introduction

Time is a fundamental stimulus dimension of input to all sensory modalities. Manipulating the temporal characteristics of a stimulus in one modality can affect time perception in other modalities, causing a discrepancy between physical stimulus timing and its perception (cf. (Eagleman, 2008) for a review). For example, changing the physical flutter rate of a clicking sound changes the apparent flicker rate of a flashing light

(e.g. (Shipley, 1964; Wada et al., 2003)). During the past decade it has been shown that the perceived number of visual events in a rapid sequence can be biased towards the number of co-occurring sounds (Shams et al., 2000), that the timing of a static sound can determine the perceived direction of visual apparent motion (Freeman & Driver, 2008) and that the perceived temporal closeness of visual events can be biased by temporally shifted auditory events (Burr et al., 2009).

More specifically, several studies have described cross-modal effects on subjective duration perception (e.g., (Chen & Yeh, 2009; Donovan et al., 2004; Klink et al., 2011; Walker & Scott, 1981), see Chapter 1.2.1.3 for a detailed review). In summary, duration judgments for audiovisual stimuli are biased towards the perceived duration of their auditory component, i.e. audiovisual duration perception seems dominated by audition. For example, the perceived duration of visual flashes can be biased by the duration of co-occurring beeps (Donovan et al., 2004; Klink et al., 2011). Participants report brief flashes that are accompanied by a longer or shorter sound to have a different duration from a flash that actually has the same duration (Donovan et al., 2004). Furthermore, flashes accompanied by a longer lasting sound are reported to last longer than their unimodal counterpart (and vice versa for shorter sounds; (Klink et al., 2011)). In this chapter I will present three experiments to further explore this phenomenon.

3.2 Experiment 1

3.2.1 Introduction

In the first experiment I aimed to replicate the effect of auditory stimuli on perceived visual duration and test whether they extend to sensitivity *enhancement* for visual duration judgments. Previous studies only investigated the (objectively detrimental) effect of incongruent auditory stimuli on visual duration judgments. Typically, however, ecologically valid audiovisual stimuli have congruent durations. Thus, auditory dominance in duration judgments could be advantageous, given the typically higher temporal resolution of the auditory system (e.g. (Chen & Yeh, 2009)).

In this experiment I asked participants to judge which of two brief flashes lasted longer. Stimulus durations were adjusted to a standard of 55 ms vs. the individual threshold for unimodal duration discrimination. I then paired the flashes to compare with sounds of congruent or incongruent (swapped) durations and tested duration judgment sensitivity for all conditions.

3.2.2 Methods

3.2.2.1 Design

On each trial subjects were presented with two visual stimuli in succession (**Figure 3-1**) and were asked to indicate which one lasted longer. There were two possible durations for the flashes, both used on every single trial, so that either the first or the second flash actually lasted longer. In

audiovisual trials the flashes were accompanied by sounds. For the first two conditions there were beeps of two durations in each trial. These durations were equal to the two durations used for flashes. However, the order of beep durations in a trial could match the order of flash durations (congruent condition) or be reversed relative to the order of flash durations (incongruent condition). In a third condition, both sounds in a trial had equal duration and matched one of the two flash durations (i.e. both sounds were brief or both sounds lasted for the longer duration). Finally, there was a unimodal, visual-only condition. In Experiment 1a the onsets of sounds and flashes was simultaneous (**Figure 3-1**). Data were analyzed in terms of signal-detection theory.

To determine whether any influence of auditory durations on visual duration judgments depended on synchronous onsets of the multisensory events, in a set of control conditions (Experiment 1b) I misaligned the onsets of auditory and visual events (by 500 ms). If the impact of auditory durations on visual duration perceptions reflects multisensory binding, it should be eliminated or reduced in the asynchronous condition (Colonius & Diederich, 2011; Meredith, Nemitz, & Stein, 1987). If instead the expected effect depended on a response bias (along the lines of a 'blind observer', simply ignoring the instruction to judge the duration of visual stimuli and instead judging the auditory durations) the effect should remain the same in the new asynchronous condition of Experiment 1b.

3.2.2.2 Participants

Seventeen healthy participants with a mean age of 26.29 (range 19–35) took part in the first experiment, nine of them female, one left-handed.

All reported normal or corrected visual acuity and normal hearing. All participants gave written informed consent in accord with University College London ethics approval. They were naïve to the purpose of the study and were paid for their time. One participant was excluded because she showed an inconsistent pattern for the unimodal threshold trials (see below). Two more participants were excluded because their performance in the visual-only condition was at ceiling in the main experiment, leaving 14 participants in the sample.

3.2.2.3 Stimuli and apparatus

Stimuli were presented on a 21′ CRT display (Sony GDM-F520) in a darkened room (screen resolution 1600 × 1200; refresh rate 85 Hz.). Participants sat with their head in a chin rest at 65 cm viewing distance. Two small stereo PC speakers were placed just in front of the monitor immediately on either side of it. Stimulus control and data recording were implemented on a standard PC, running E-Prime 2 Professional (Psychology Software tools, Inc., www.pstnet.com). Unspeeded manual two-choice responses were made using a standard PC keyboard.

Each visual stimulus comprised a white disk extending 1.2° in visual angle with its midpoint at 3° below a central fixation cross on a gray background. On each trial a pair of disks was flashing consecutively with varying durations from 55 to 165 ms (see below).

The auditory stimulus was a 900-Hz pure tone sampled at 44.100 kHz with durations also varying from 55 to 165 ms (see below). Sound level was measured with an audiometer and set to ~70 dB(A).

3.2.2.4 Procedure

3.2.2.4.1 Unimodal threshold determination

Only visual stimuli were presented during this part of the experiment. On each trial participants were presented with a pair of disks flashing in two consecutive time windows separated by a stimulus onset asynchrony (SOA) of 1000 ms. While one of the two visual stimuli had a constant duration of 55 ms (the standard), the other lasted slightly longer, with its duration varying between 66 and 165 ms (10 possible incremental steps of one frame at 85 Hz, i.e., ~11 ms). The latter stimulus type will be referred to as the "longer" stimulus. Each of the resulting 10 pairings of standard and longer stimuli was repeated 10 times per block. Each participant completed two to three blocks. The pairwise order of standard and longer stimuli was counterbalanced between trials, with standard-longer or longer-standard pairwise sequences being equiprobable.

On each trial participants were instructed to indicate whether the first or the second flash lasted longer, by pressing a corresponding button on the keyboard ("1" or "2"). This way I could identify the visual duration discrimination threshold for each participant individually. Threshold was defined as corresponding to the increase in duration for the longer stimulus whose duration allowed correct identification of it as longer in ~75% of cases. Participants were able to discriminate differences of durations correctly in 73.78% of cases for those stimulus pairings containing the longer stimulus that was identified as threshold (standard error (SE): 1.6%).

The average durations of the longer stimuli identified as threshold were 103.4 (±3.97 SE) ms.

3.2.2.4.2 Main experiment

In each trial of the main experiment, participants were presented with the pair of visual stimuli previously identified as around threshold. Again, the order of standard and longer visual stimulus was counterbalanced and equiprobable, with participants again asked to indicate which of the two consecutive flashes lasted longer. But the main experiment now consisted of 10 conditions (5 in Experiment 1a, and 5 in Experiment 1b – note that all 10 conditions were intermingled but are presented separately here for ease of exposition). These 10 conditions differed with regard to whether, when, and how any sounds were presented with the flashes. Participants were instructed to ignore all sounds played during the experiment and to judge only the duration of the visual stimuli.

Two pure tone durations were selected for each participant – one lasting 55 ms and thus matching the standard visual stimulus in duration, the other auditory duration matching the participant-specific longer visual stimulus identified as threshold during the preceding visual titration task. These two pure tones were then combined with the flashes according to condition. There were two main classes of conditions: potentially synchronous (Experiment 1a) or asynchronous (Experiment 1b). In the potentially synchronous conditions, tone onset was temporally aligned with the visual onsets; whereas in the potentially asynchronous conditions, the onset of tones was delayed for 500 ms (thus 180° out of "phase" if one

considers the pair of visual stimuli as a cycle, for which 180° yields the maximum possible phase offset) relative to flash onsets. In either of the potentially synchronous or asynchronous situations, there were five possible conditions: audio–visual congruent (same order of durations in the two modalities), audio–visual incongruent (opposite orders of durations in the two modalities), both-long auditory stimuli, both-short auditory stimuli, or a purely visual condition (c.f. **Figure 3-1**). The purely visual condition was of course actually equivalent for "synchronous" and "asynchronous" conditions. Trials from these conditions were randomly split in half and assigned to the synchronous or asynchronous conditions for each participant, thus allowing for a 5 × 2 factorial analysis of variance (ANOVA) of the data; see below.

Each block contained 10 repetitions for each of the 10 conditions in a randomized order. Every participant repeated three to four of these blocks.

3.2.2.5 Data analysis

I computed sensitivity (d') for the duration discrimination task, for each participant and condition using the standard formula (Macmillan & Creelman, 2004):

$$d' = \mathrm{z(H)} - \mathrm{z(F)}$$

where z(H) stands for the z-transform of the hit rate, while z(F) stands for the z-transform of the false-alarm rate. To address extreme cases

(where false rates were zero) I adjusted all d' values as suggested in (Snodgrass & Corwin, 1988): false alarm rates were calculated as the number of false alarms +0.5, divided by the number of no-signal trials plus one (and, equally, hit rates as the number of hits + 0.5, divided by the number of signal trials plus one; c.f. (Macmillan & Creelman, 2004)).

d' values were analyzed using repeated-measures analysis of variance (ANOVA), with SYNCHRONY (synchronous/asynchronous) and audio–visual CONDITION (Congruent; Incongruent; Short sounds; Long sounds, Purely visual) as within-subjects factors; followed up by pairwise *t* -tests where appropriate.

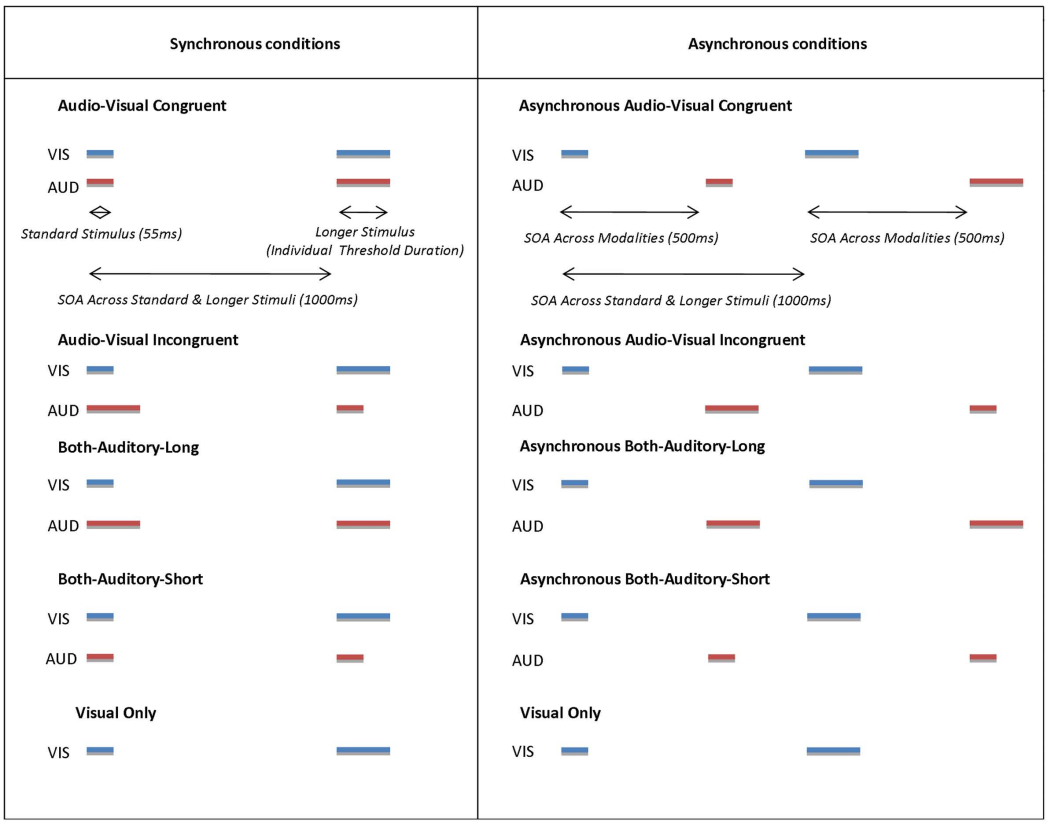

Figure 3-1 Schematic timelines representing conditions in Experiment 1A and 1B. In Experiment 1A all corresponding auditory (red lines marked with label "AUD") and visual (blue lines marked with label "VIS") stimuli had the same onset (synchronous conditions) while in Experiment 1B visual stimuli always preceded auditory stimuli by 500 ms (asynchronous conditions). In both situations, each pair of stimuli within one modality was separated by a 1000-ms interval; the order of short and long stimuli could be the same across auditory and visual modalities (congruent) or reversed between them (incongruent). In the both-auditory-short condition, both successive sounds had the shorter visual duration (and vice versa for both-auditory-long). Finally a visual-only condition served as baseline measure.

3.2.3 Results

The sensitivity (d') results are shown in **Figure 3-2 A** (synchronous conditions) and **Figure 3-2 B** (asynchronous conditions), as group means with SE. Note the higher sensitivity specifically in the synchronous audio–visual congruent condition (**Figure 3-2 A**), leftmost bar). The overall 5 × 2

ANOVA showed a main effect of SYNCHRONY [F (1,13) = 12.09, p < 0.01], a significant main effect of audio–visual CONDITION [F (4,52) = 6.68, p < 0.001] and critically a significant interaction between these two factors [F (4,52) = 5.96, p < 0.001].

To identify the source of the interaction, first two separate one-way ANOVAs were performed for synchronous or asynchronous datasets, with the five-level factor of condition. While the asynchronous conditions did not differ significantly from each other [F (4,52) = 0.56, p = 0.69 for the main effect], the synchronous conditions did [F (4,52) = 10.91, p < 10^-5]. Exploratory pair-wise t-tests for the asynchronous conditions confirmed no significant differences between any of these conditions (all p > 0.20). Pairwise t -tests for the synchronous conditions showed that sensitivity in the synchronous audio–visual congruent condition (d' = 1.93 ± 0.23 SE) was significantly higher than in all the other conditions, as follows: (i) versus the synchronous audio–visual incongruent condition [d' = 0.20 ± 0.15 SE; t (13) = 4.93; p < 0.001]; (ii) versus the both-auditory-short condition [d' = 1.07 ± 0.16 SE; t (13) = 3.63; p < 0.01]; (iii) versus the both-auditory-long condition [d' =1.11±0.19SE; t(13)=3.0, p=0.01].

When compared to the visual-only baseline measure (d' = 1.01 ± 0.18 SE), I found that: (i) visual duration discrimination was significantly enhanced in the synchronous audio–visual congruent condition [t (13) = −3.38; p < 0.01]; (ii) was significantly impaired in the synchronous audio–visual incongruent condition [t (13) = 3.44, p < 0.01]; (iii) was not significantly affected in the both-auditory-long or short conditions (all p >

0.71).

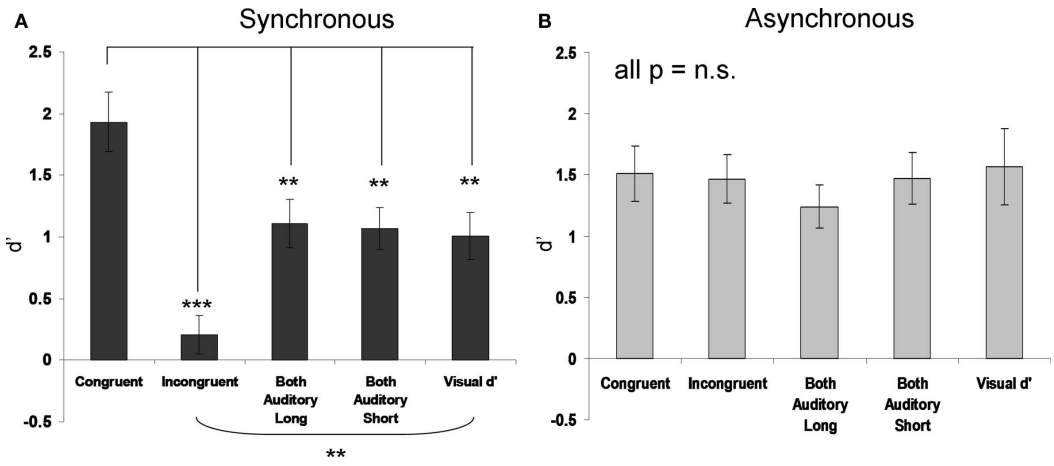

Figure 3-2 Mean visual duration discrimination sensitivity (d′) for each condition in Experiment 1A (A) and 1B (B). Asterisks above bars in **A** indicate significant differences relative to the synchronous–congruent condition that gave best performance, see leftmost bar in left graph. Asterisks below bars indicate a significant difference between the audio-visual incongruent and visual-only conditions (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$). None of the pairwise contrasts were significant for the asynchronous conditions (see right graph). Error bars indicate +/- one standard error of the mean (s.e.m.).

3.2.4 Discussion

In Experiment 1 I replicated the finding that duration discrimination for visual stimuli is modulated by the duration of co-occurring auditory stimuli. Further, I extended this finding by showing that it also applies to congruent auditory stimuli. Congruent auditory stimulation can enhance audiovisual duration sensitivity in much the same way that incongruent auditory stimuli can decrease performance. Furthermore, this was only the

case when those auditory stimuli were synchronous with the visual events, rather than being delayed by 500 ms. This elimination of the effect for the asynchronous case renders a genuine multisensory effect likely and speaks against a mere 'blind observer' response bias (see above).

The second experiment was designed to further probe the multisensory nature of the effect. If the observed effect is indeed a multisensory one (rather than a response bias towards the auditory modality), there presumably is a limit to how much auditory and visual durations can differ in order to be bound into a single percept. If the auditory event would endure much longer than the corresponding visual event, it should become less plausible that they arise from the same cause. Consequently the auditory influence should begin to wane. This is what I tested in Experiment 2.

3.3 Experiment 2

3.3.1 Introduction

The second experiment repeated those conditions of Experiment 1 for which effects had been observed (i.e., synchronous–congruent, synchronous–incongruent and visual-only baseline). However, now either the sounds had the same possible durations as the flashes, or else the sounds were tripled in duration such that they would no longer plausibly correspond to the flashes. A mere response bias, akin to a blind observer simply reporting the durations of the sounds, should lead to a similar outcome in either case; a multisensory effect on the other hand is expected to be reduced (or even disappear) for tripled sound durations.

3.3.2 Methods

3.3.2.1 Participants

Fifteen new healthy participants with a mean age of 24.53 (range 18–34) took part, seven females, one left-handed, all reporting normal or corrected visual acuity, and normal hearing.

3.3.2.2 Stimuli and apparatus

The setup was as for Experiment 1, but some conditions were dropped and two new ones added. I repeated the synchronous–congruent, synchronous-incongruent, and visual-only conditions. The two new conditions were versions of the synchronous–congruent and synchronous–incongruent conditions with tripled auditory durations. In these tripled conditions, each sound still had a synchronous onset with each flash, but the sounds now lasted three times as long, so they had a much later offset than the visual events.

3.3.2.3 Procedure

3.3.2.3.1 Unimodal threshold determination

This aspect of the procedure was the same as for Experiment 1. On average, participants were able to discriminate durations correctly for 76.33 (±1.5 SE)% of cases for those stimulus pairings that contained the longer visual stimulus identified as threshold. The average duration of the longer

visual stimuli identified as thresh- old was 94.5 (±3.49 SE) ms, so 49.5 ms longer than the standard stimulus, with the visual disks used.

3.3.2.3.2 Main experiment

The procedure resembled Experiment 1, but with only five conditions, two of which were new (see above).

3.3.2.4 Data analysis

For each participant and condition we computed sensitivity (d') for each stimulus condition as in Experiment 1. d' values for the four audio–visual conditions were analyzed using a repeated-measures two-way ANOVA, with SOUND LENGTH (tripled or untripled) and audio–visual duration CONGRUENCY (congruent versus incongruent) as factors. In addition pairwise *t*-tests were used to compare performance in the purely visual baseline against the remaining conditions.

3.3.3 Results

The sensitivity (d') results are shown in **Figure 3-3**, as group means with SE. Note the higher sensitivity in the audio–visual untripled synchronous–congruent condition (leftmost bar), replicating the effects obtained in Experiment 1a (c.f. **Figure 3-2 A**).

The two-way ANOVA showed no main effect of SOUND LENGTH [$F(1,14) = 1.94$, $p = 0.18$], a significant main effect of CONGRUENCY [$F(1,14) =$

49.33, $p < 0.00001$] and a significant interaction between these two factors [$F(1,14) = 10.99$, $p < 0.01$], because the congruent/incongruent difference was larger in the untripled than tripled case. Sensitivity in the audio–visual untripled synchronous–congruent condition ($d' = 2.07 \pm 0.16$ SE) was significantly higher than in all the other conditions, as follows: (i) versus the visual-only duration condition [$d' = 1.58 \pm 0.16$ SE; $t(14) = 2.86$; $p = 0.013$]; (ii) versus the tripled congruent condition [$d' = 1.72 \pm 0.18$ SE; $t(14) = 2.82$; $p = 0.014$]; (iii) versus the untripled incongruent condition [$d' = 0.11 \pm 0.26$ SE; $t(14) = 8.85$; $p < 10^{-6}$]; (iv) versus the tripled incongruent condition [$d' = 0.82 \pm 0.25$ SE; $t(14) = 4.8$, $p < 0.001$].

Thus, by prolonging the duration of the auditory stimuli to triple that of the visual stimuli, the significant enhancement (relative to visual-only baseline) obtained for the congruent audio– visual durations was abolished. At the same time triple-duration auditory stimuli still produced some sensitivity decrease for incongruent stimuli. However this remaining decrease was significantly reduced for the tripled versus untripled case [$t(14) = -2.69$, $p = 0.018$].

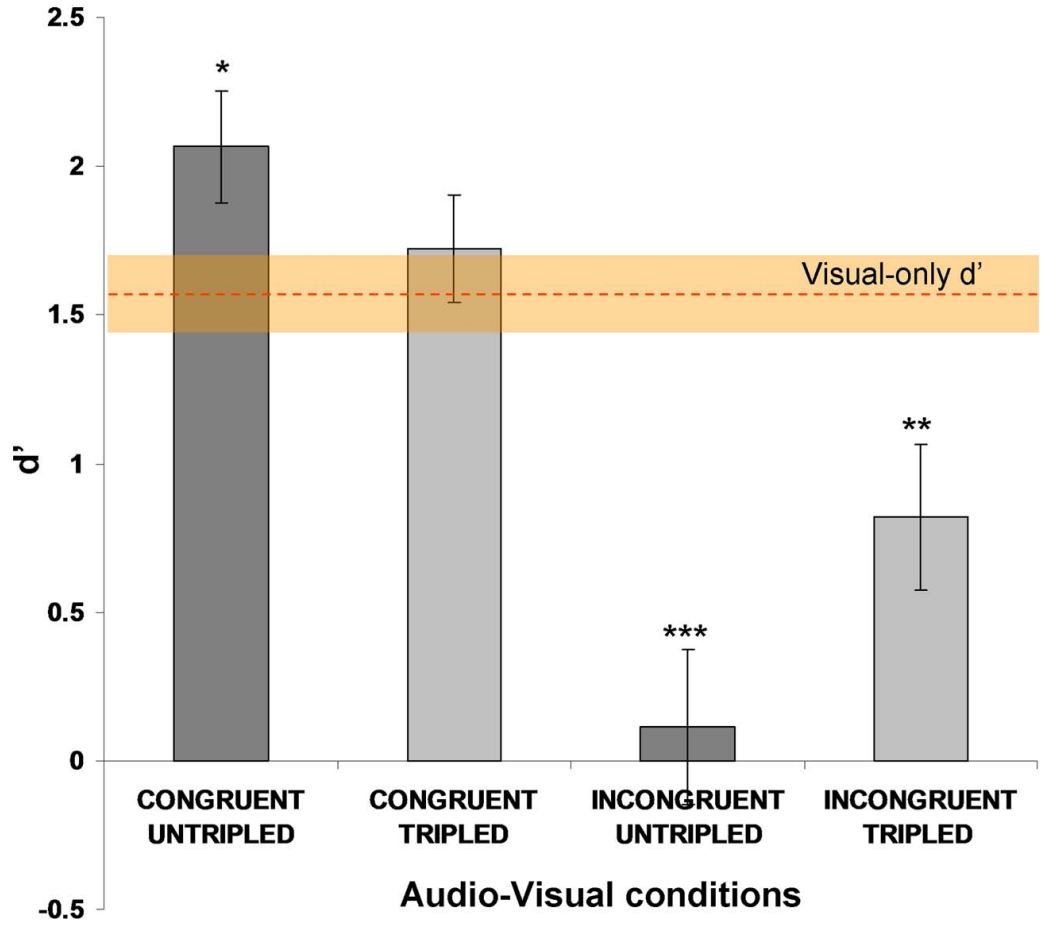

Figure 3-3 Mean visual duration discrimination sensitivity (d′) for each condition in Experiment 2. Error bars indicate +/- one SEM. Asterisks above bars point to significant differences relative to the visual-only baseline, with the latter represented here by the orange dashed line with SEM shading. The significant enhancement or decrease of sensitivity, for the untripled congruent and incongruent conditions (respectively) replicates the findings of Experiment 1a. These effects were eliminated or reduced (respectively) for the corresponding two new tripled conditions, in which sounds and flashes still had synchronous onsets but sounds lasted three times as long.

3.3.4 Discussion

In Experiment 2 I replicated the main results of Experiment 1a, showing significant enhancement of visual duration discrimination

sensitivity (d') by congruent auditory stimuli and a significant decrease of sensitivity for incongruent stimuli. Additionally, Experiment 2 showed that the sensitivity enhancement for congruent stimuli relative to the visual-only baseline was abolished for prolonged auditory stimuli. Some sensitivity decrease for incongruent audio–visual pairings remained but at a significantly reduced level in the tripled-sound-duration condition. These results show that the impact of auditory durations on visual duration discrimination is larger when the sounds and flashes endure for a similar order of magnitude. They thus support the interpretation of the effect as multisensory rather than a mere response bias (just as the findings of Experiment 1 did, see above).

However, the nature of such a multisensory effect is still unclear. It could take place at the level of a supramodal general duration perception mechanism, like an internal 'clock' (c.f. above and General Introduction). Alternatively, it is conceivable that auditory stimuli can alter the actual duration of short visual percepts (Berger et al., 2003) and the corresponding neural activity in the visual system. If longer lasting sounds can indeed sustain the perception of co-occurring short visual events, this should enhance visual sensitivity for orthogonal non-temporal stimulus qualities as well. This prediction was tested in Experiment 3.

3.4 Experiment 3

3.4.1 Introduction

The aim of Experiment 3 was to test whether pairing a visual stimulus with a longer lasting sound would yield prolonged perception of the visual stimulus. If so, sensitivity for the visual stimulus should be facilitated for longer sounds, similar to the visual detection improvement expected for a genuinely longer lasting visual stimulus.

Furthermore I anticipated any such effect to be restricted to a critical time window of audiovisual integration. Previous studies point to the importance of cross-modal stimulus onsets falling within a time window of about 100 ms for audiovisual binding to occur (Bolognini, Frassinetti, Serino, & Làdavas, 2005; Romei, Murray, Merabet, & Thut, 2007). The results of Experiment 2 pointed to a similar time window regarding the effect of prolonging sounds on visual stimulus duration judgments. If the duration of sounds was stretched too far beyond the visual stimulus offset, the effect disappeared. I therefore aimed to parametrically vary the duration of co-occurring sounds up to about 100 ms and add an additional data point for a sound duration presumed to fall well beyond this temporal window of integration. My hypothesis was that if there was an effect of sound duration on visual sensitivity, sensitivity would continuously increase with sound duration but fall back to baseline level for the longest sound duration purposefully chosen to fall beyond the window of audiovisual integration.

3.4.2 Methods

3.4.2.1 Participants

Twenty-eight healthy participants were recruited for this experiment (mean age 25.1 years, range 25-30 years; 19 females, all right handed). All reported normal or corrected visual acuity and normal hearing. All participants were paid for their time.

All participants gave written informed consent to take part in this study, according to the Declaration of Helsinki. The study was approved by the UCL Research Ethics Committee.

3.4.2.2 Stimuli and apparatus

The setup used was the same as for Experiments 1 and 2. Stimulus control and data recording were implemented on a standard PC, running a MATLAB script using functions of Psychophysics Toolbox 3 (Kleiner, Brainard, & Pelli, 2007).

In each trial, a rectangle containing dynamic white noise (mean luminance: 4.8 cd/m^2; size: 23.5 x 17.7 degrees of visual angle) was presented for two consecutive intervals, each lasting 1059 ms (i.e. 90 frames at a video refresh rate of 85Hz), with an SOA of 300 ms between the two displays. A fixation dot extending 0.22 degrees in visual angle was superimposed at the middle of the noise rectangle, which was centered on the screen. The fixation dot was visible throughout the whole experiment, and changed its color from red to green as a 'go' signal for responses in between trials. The target visual stimulus was a transparent Gabor patch (alpha blending factor of .6) which was briefly flashed at 353 ms after the

onset of the first or second dynamic noise rectangle. The Gabor patch was composed of a 2D sinusoidal luminance grating with spatial frequency of 2.69 cycles per degree visual angle within a Gaussian amplitude envelope with a standard deviation of 10 pixels. It was embedded in the white noise visual stimulation with its center position 1.4 degrees visual angle below the fixation dot. The luminance amplitude of the Gabor patch was set to individual threshold and its duration varied with experimental conditions (see below).

The auditory stimulus was a 400 Hz sinusoidal pure tone sampled at 44.1 kHz with 8 different durations (~24, ~36, ~48, ~60, ~72, ~84, ~96 and ~190 ms). Sound level was set to a ~70 dB(A). In the audiovisual trials of the main experiment, sound stimuli of equal duration were presented at 353 ms after the onset of both of the white noise rectangles.

3.4.2.3 Procedure

3.4.2.3.1 Unimodal threshold determination

Only visual stimuli were presented during this part of the experiment. In each trial a Gabor patch of ~24ms duration was embedded in one of two consecutive white noise displays with its midpoint at 1.4 degrees below fixation. In between trials the red fixation dot turned green, indicating participants should respond as to whether the Gabor patch was presented in the first or second white noise interval by pressing “1” or “2” on a keyboard. Across trials we pseudo-randomly varied the luminance amplitude of Gabor patches following a constant stimuli design (8 steps within a range of peak

luminance measurements between 4.8 cd/m^2 to 6.3 cd/m^2). This allowed me to identify the luminance threshold for each participant individually. Participants completed 2 blocks, each containing 14 trials of each of the 8 luminance amplitudes tested, i.e. 112 trials in total.

The threshold was defined as the luminance amplitude allowing participants to correctly answer in 60% of the cases. In order to determine threshold luminance, a sigmoid function was fitted to the visual titration data of each individual.

3.4.2.3.2 Main experiment

In the main experiment, participants were again presented with a consecutive pair of dynamic white noise rectangles. Again, a Gabor patch was embedded in either the first or second white noise interval, and participants had to indicate whether they saw the flash in the first or second interval after each trial. The luminance amplitude of the target Gabor patch was set to a fixed value corresponding to the individual threshold, determined by the unimodal threshold procedure for each participant. Trials in the main experiment fell in two conditions in pseudo-random order. In *visual* trials the flashing Gabor patch lasted for one of eight different durations (~24, ~36, ~48, ~60, ~72, ~84, ~96, ~190 ms), varying pseudo-randomly between trials. In *audiovisual* trials, the Gabor patch flash duration was fixed at ~24ms. While in visual trials no sound occurred, audiovisual trials additionally contained a pure tone auditory stimulus which was played twice with onsets at 353 ms after the first and second

white noise interval respectively (i.e. synchronous with the Gabor patch onset in the target interval and at the matching time point during the non-target interval). The duration of tones pseudo-randomly varied between trials, corresponding to the same eight durations that the flashes could have in the unimodal visual condition (~24, ~36, ~48, ~60, ~72, ~84, ~96 and ~190 ms). Tone durations were always equal for the first and second interval of a given trial. Participants were instructed that auditory stimuli were irrelevant for the purpose of the task and therefore to ignore them. Participants completed 6 blocks of 80 trials each for a total of 30 stimuli per duration tested. The procedure is illustrated in **Figure 3-4**.

3.4.2.4 Data analysis

For each participant I computed visual sensitivity (d') for the visual detection task independently for each of the visual and corresponding auditory-visual durations. This was done as for Experiments 1 and 2 (see above).

d' was analyzed using repeated-measure analysis of variance (ANOVA), with CONDITION (visual only and audiovisual) and DURATION (~24, ~36, ~48, ~60, ~72, ~84, ~96 and ~190 ms) as within subjects factors followed up by paired t-tests where appropriate.

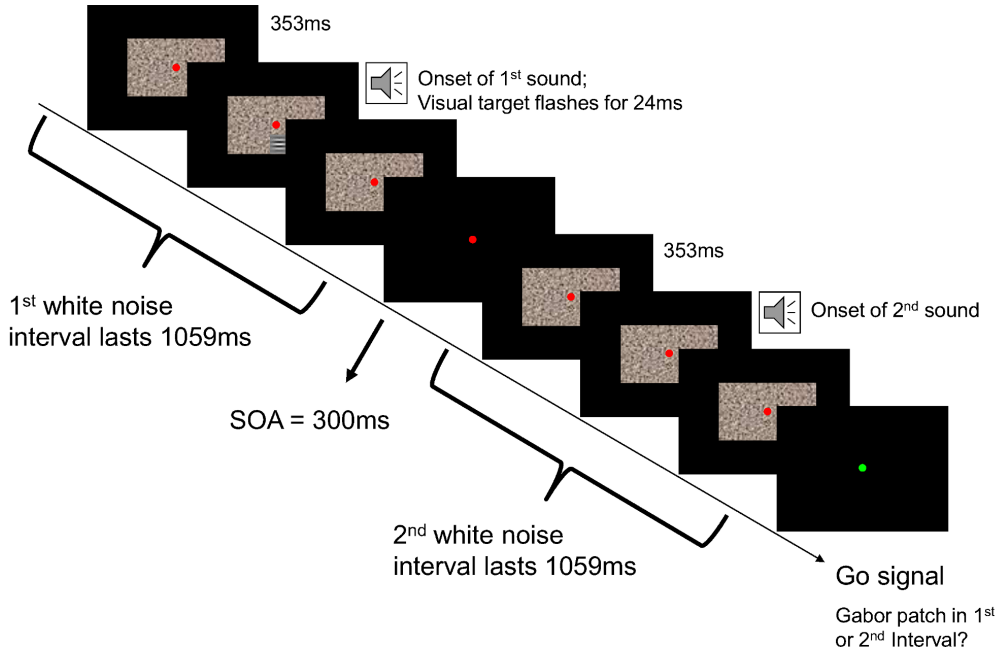

Figure 3-4 Illustration of a trial in the main experiment. Participants fixated a central red dot, while two consecutive intervals of dynamic white noise were presented on the screen. In either the first or second interval a Gabor patch was flashed for ~24 ms at 1.4 degrees visual angle below fixation. The target flash appeared equiprobably in the first and second interval (each interval lasting 1059 ms, with the onset of the Gabor flash at 353 ms). In the example depicted the target flash appears in the first interval. Additionally, in both intervals a sound of variable duration (~24 to ~190 ms) was presented (sound onset 353 ms after interval onset for both intervals). In trials of a second, visual only, condition no sounds were played and the duration of Gabor patch flashes was variable (~24 to ~190 ms, matching sound durations in the audiovisual condition). After stimulus presentation the fixation dot turned green, indicating participants should report whether they perceived the Gabor patch in the first or second interval by button press.

3.4.3 Results

3.4.3.1 Effects of visual and auditory stimulus duration on visual sensitivity

The sensitivity (d') group means and standard errors are shown in **Figure 3-5 A** (visual stimuli alone) and **Figure 3-5 B** (audio-visual stimuli). Note the increase in sensitivity in **Figure 3-5 A** as a function of visual stimulus duration and the corresponding increase in sensitivity in **Figure 3-5 B** for auditory stimuli of ~60, ~72, ~84 and ~96 ms duration before falling back towards baseline level at ~190 ms. Please refer to **Table 3-1** and **Table 3-2** for the complete set of signal detection results.

The 2 x 8 ANOVA showed a main effect of CONDITION ($F(1,27)=563.73$; $p<0.000$), a main effect of DURATION ($F(7,189)=60.7$; $p<0.000$) and a significant interaction between these two factors ($F(7,189)=46.92$; $p<0.000$). I broke down our analysis by the factor Condition, thus testing the factor Duration for visual only trials and audiovisual trials separately. For visual only trials I found a significant effect of DURATION ($F(7,189)=92.17$; $p<0.0000$). Paired t-tests showed that compared to the shortest visual stimulus duration (~24 ms, the baseline measure: BSL), all other visual stimulus durations enhanced visual sensitivity (all *p*s<0.004, Bonferroni corrected). Crucially, also the repeated-measure ANOVA for audiovisual trials showed a significant effect of auditory stimulus Duration on visual sensitivity ($F(7,189)=2.29$; $p=0.029$).

Paired t-tests showed that compared to the shortest, baseline audiovisual stimulus (~24 ms), visual sensitivity was enhanced for auditory stimulus durations of ~60 ms ($t(27)=3.08$; $p=0.03$, Bonferroni corrected), ~72 ms ($t(27)=3.75$; $p=0.006$, Bonferroni corrected) ~84 ms (($t(27)=4.06$;

p=0.005, Bonferroni corrected) and ~96 ms (t(27)=3.84; p=0.005, Bonferroni corrected). All other auditory durations (~36, ~48 and ~190 ms) did not significantly differ from our BSL (all *ps*>0.23, Bonferroni corrected). The maximum sensitivity in the audiovisual condition (~.6 d') was lower than in the visual condition (~3 d') and similarly the maximum enhancement relative to the respective baseline was smaller in the audiovisual (~7% increase in accuracy) than the visual condition (~35% in accuracy). Furthermore, sensitivity for the shortest (~24 ms) audiovisual stimulus was significantly lower than for the shortest (~24 ms) unimodal visual stimulus (t(27)=-3.36, p<.05, Bonferroni corrected), while sensitivity for all other sound durations was not significantly different from sensitivity for the shortest unimodal visual stimulus (all p values >.24).

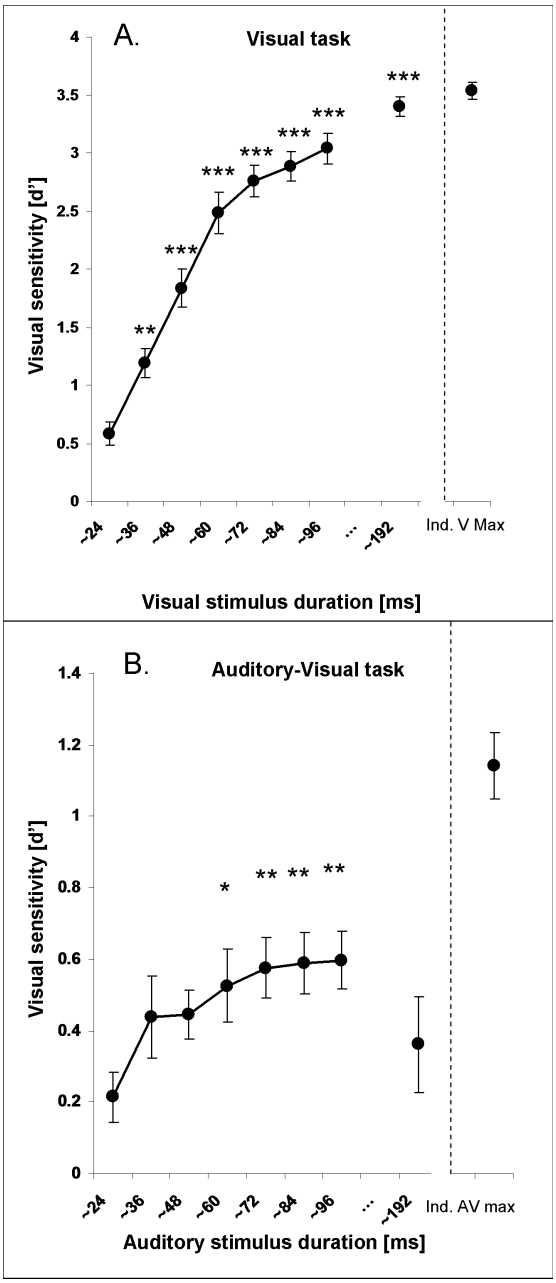

Figure 3-5 Effect of stimulus durations on visual sensitivity. Mean visual discrimination sensitivity (d', SEM indicated) for varying visual stimulus durations (**A**), upper panel) and for visual stimuli of fixed duration (~24 ms), paired with auditory stimuli of varying durations (**B**), lower panel). Asterisks indicate significant enhancement in visual sensitivity relative to the shortest (audio-)visual stimulus (leftmost data point in **A** and **B** (* p<.05, ** p<.01 ***p<.001; all Bonferroni corrected). The rightmost data point represents the average maximum of the visual sensitivity enhancement across participants in the visual task ('Ind. V max', **A**) and audiovisual task ('Ind. AV max', **B**).

Table 3-1 Hit rates (HIT), false alarm rates (FA) and criteria (c) for the visual condition. Cells contain the mean and standard error of the mean (S.E.M.) across participants. Signal trials were defined as the ones in which the visual stimulus was displayed during the first interval.

Duration	HIT % (S.E.M.)	FA % (S.E.M.)	c (S.E.M.)
24ms	70% (± 3%)	48% (± 3%)	-0,21 (± 0,07)
36ms	76% (± 3%)	34% (± 3%)	-0,15 (± 0,07)
48ms	84% (± 3%)	23% (± 3%)	-0,16 (± 0,07)
60ms	88% (± 2%)	11% (± 2%)	0,009 (± 0,06)
72ms	94% (± 1%)	9% (± 2%)	-0,06 (± 0,05)
84ms	94% (± 1%)	6% (± 2%)	0,01 (± 0,04)
96ms	96% (± 1%)	6% (± 2%)	-0,05 (± 0,04)
190ms	99% (± 1%)	3% (± 1%)	-0,05 (± 0,05)

Table 3-2 Hit rates (HIT), false alarm rates (FA) and criteria (c) for the audiovisual condition. Cells contain the mean and standard error of the mean (S.E.M.) across participants. Signal trials were defined as the ones in which the visual stimulus was displayed during the first interval.

Duration	HIT % (S.E.M.)	FA % (S.E.M.)	c (S.E.M.)
24ms	64% (± 2%)	53% (± 3%)	-0,21 (± 0,07)
36ms	65% (± 4%)	50% (± 3%)	-0,21 (± 0,08)
48ms	62% (± 3%)	42% (± 3%)	-0,06 (± 0,07)
60ms	65% (± 3%)	44% (± 3%)	-0,13 (± 0,05)
72ms	64% (± 4%)	41% (± 3%)	-0,06 (± 0,08)
84ms	68% (± 3%)	45% (± 3%)	-0,18 (± 0,06)
96ms	65% (± 3%)	40% (± 3%)	-0,06 (± 0,06)
190ms	62% (± 4%)	49% (± 3%)	-0,15 (± 0,08)

3.4.3.2 Correlation between effects of visual and auditory stimulus duration

The magnitude of visual sensitivity enhancement for prolonged sounds relative to the shortest sound (i.e. BSL corrected values) varied considerably across participants (range: 0 to 2.35 d', mean .93 d'; SD: .51 d'). The same was true with regard to the magnitude of enhancement for genuinely prolonged visual stimuli relative to the shortest visual stimulus, i.e. BSL corrected values (range: 0.86 to 3.84 d', mean: 2.95 d'; SD: .69 d'). Thus there were individual differences with regard to both processes: visual sensitivity enhancement by prolonging visual stimulus duration and by prolonging the duration of co-occurring sounds (for individual differences in audiovisual integration cf. (Nath & Beauchamp, 2012; Spence & Squire, 2003; Stone et al., 2001); Chapter 4). I hypothesized that the size of these effects would be correlated across participants if they stem from similar neural mechanisms.

Furthermore, participants also differed with regard to the particular sound duration for which they showed maximum visual sensitivity enhancement. I speculated that this might reflect genuine trait-like differences between participants, such as the individual width of the multisensory window of integration (cf. (Spence & Squire, 2003; Stone et al., 2001)). Based on these assumptions I hypothesized that the maximum visual sensitivity enhancement (relative to BSL) for a given participant would be the best indicator for this participant's effect size relative to other participants. The effect size in question is the *individual* peak auditory

enhancement that I will refer to as 'the maximum audiovisual sensitivity enhancement' (Ind. AV max).

In a similar way I calculated the *individual* 'maximum visual sensitivity enhancement' (Ind. V max) in the unimodal condition. Note that the *absolute* size of both 'Ind. V max' and 'Ind. AV max' are likely to be inflated and thus non-informative per se. Still they appear as the 'fairest' way of quantifying *relative* individual effect sizes without biasing towards a particular width for the window of integration.

I therefore tested the correlation between 'Ind. AV max' and 'Ind. V max'. The individual maxima in duration induced enhancement were significantly correlated between both conditions (r=.38, p<.05; see **Figure 3-6**).

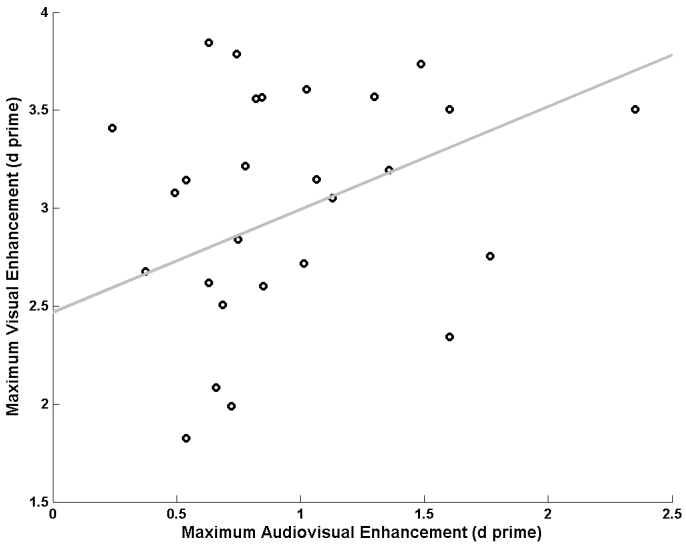

Figure 3-6 Correlation between visual and audio-visual enhancement. Correlation between individual peak auditory-induced enhancement ('Ind. AV max') and peak visually-induced enhancement ('Ind. V max'). Note that the maximum effective auditory and visual durations varied between participants and were thus determined on an individual basis (c.f. Methods for details).

3.4.4 Discussion

The results of experiment three indicate that visual sensitivity depends on the duration of co-occurring auditory stimuli. Specifically, visual detection sensitivity (d') for a ~24 ms visual flash was significantly enhanced for auditory stimuli whose durations were between ~60 and ~96 ms, as compared to performance at baseline (matching auditory duration of ~24 ms). However, no such visual sensitivity enhancement was found for an auditory stimulus lasting much longer than this critical time window (~190 ms). A rather surprising aspect of the results is that the baseline level of performance in the audiovisual condition was significantly lower than the unimodal baseline performance. Participants' detection performance became significantly worse when a ~24 ms flash was accompanied by an auditory stimulus of matching duration (as compared to no accompanying sound).

These results are consistent with the finding of Experiments 1 and 2, showing that auditory stimuli can bias duration judgments for co-occurring visual stimuli (c.f. (Chen and Yeh, 2009; Donovan et al., 2004; Klink et al., 2011; Walker and Scott, 1981)). They go beyond these findings by showing that the duration of auditory stimuli also impacts objective visual performance for non-temporal visual stimulus qualities.

The aim of Experiment 3 was to test whether the multisensory effects of Experiment 1 and 2 reflect an effect of sustaining visual perception. If this was the case, it should affect duration judgments as well as non-temporal qualities of visual perception, including detection sensitivity for visual stimuli.

The findings support this hypothesis and characterize the enhancement of visual sensitivity for sounds of longer duration within a restricted time window. This time window (~60-96 ms) is consistent with previous findings regarding critical time windows for audiovisual integration (see e.g. (Bolognini et al., 2005; Romei et al., 2007)).

3.4.4.1 Lower baseline performance in the audiovisual condition

Despite the predicted pattern of results within the audiovisual condition, a comparison across conditions yielded a surprising result in need of explanation. Lower baseline performance in the audiovisual condition was neither predicted by my hypotheses, nor by the results of previous studies. Generally, the mere presentation of auditory stimuli during a visual task can modulate visual performance (e.g. (Bolognini et al., 2005; Cappe, Thut, Romei, & Murray, 2009; Frassinetti, Bolognini, & Làdavas, 2002; Kim, Peters, & Shams, 2011; Leo, Romei, Freeman, Ladavas, & Driver, 2011; Noesselt et al., 2010; Shams et al., 2000; Spence & Driver, 2004; Vroomen & De Gelder, 2000; Vroomen & Keetels, 2009)) as well as responses in early visual areas (Cappe, Thelen, Romei, Thut, & Murray, 2012; Cappe et al., 2010; Lakatos et al., 2009; Murray, Cappe, Romei, Martuzzi, & Thut, 2012; Romei et al., 2009, 2007; Wang, Celebrini, Trotter, &

Barone, 2008). But even studies using very similar stimuli and paradigms found a detection sensitivity *enhancement* for audiovisual vs. visual stimuli, rather than a detrimental effect, as in the results of Experiment 3 (Chen, Huang, Yeh, & Spence, 2011; Noesselt et al., 2010).

In the study by (Noesselt et al., 2010) participants had to decide in each trial, whether a Gabor patch was flashed in a cued peripheral region of interest or not. As in our experiment, flashes were quite brief (17 ms) and could be accompanied by a sound of matching duration. The presence or absence of sounds was not informative (sounds were as likely to be played in no-signal as in signal trials). Flash intensity was thresholded to 55-65% (low intensity) or 85-95% correct (high intensity) for unimodal visual stimuli. The presence of sounds yielded a significant detection sensitivity enhancement for low intensity stimuli. This condition was very similar to the experiment presented here, with the only differences being that my stimuli were embedded in dynamic noise patterns and I used a two interval forced choice design, rather than a simple detection design. The latter difference could in theory be of importance – simple detection designs are more prone to biases in decision criterion and sound induced performance enhancements could thus be due to criterion shifts. Participants in this study (Noesselt et al., 2010) indeed showed a conservative bias for low intensity stimuli, but the improvement in the audiovisual condition was in objective performance (d') and not accompanied by a shift towards a more liberal criterion. The differences in results between this study and mine are thus unlikely to derive from the differences in task design.

The design of (Chen et al., 2011) was even closer to the one presented here. Here, participants had to decide in which of two intervals a Gabor patch was flashed for 17 ms, and stimuli were embedded in dynamic noise. The only difference was that in this study stimulus frames were interleaved with frames of noise, while our stimuli were superimposed with the noise mask (see above, Methods). The authors measured stimulus intensity thresholds for fixed steps of noise intensities and compared the resulting threshold curves in the presence vs. absence of a non-informative, co-occurring sound with matching duration. The co-occurring sound led to significantly lowered detection thresholds, but this effect was restricted to one out of seven noise intensities. Further, an example set of psychometric functions provided (their Figure 2) points to the possibility that the sound-induced enhancement for stimulus intensities around threshold (i.e. 75% correct for unimodal stimuli) might be much less pronounced, or even reversed for lower stimulus intensities (yielding performance levels of 55-65%; note that this is the performance level we aimed for in our thresholding procedure). Taken together, baseline performance enhancement for audiovisual stimuli – in the particular design I used – seems to be rather subtle and dependent on specific combinations of performance level, signal intensity and noise intensity.

Still, to my knowledge, I am the first to describe a significant detrimental effect of co-occurring sounds on visual detection sensitivity. Further research is needed to investigate this effect. One might speculate that very short sound transients can have a detrimental effect on visual sensitivity due to modality specific latencies in neural processing and the

tendency of the subjective point of audiovisual synchrony to be shifted towards a visual lead (Vroomen & Keetels, 2010). Depending on the nature of cross-modal interactions, a sound-induced boost in visual neural activity might precede the onset of visually evoked activity. This in turn could lower the signal to noise ratio of visual stimulus evoked activity. However both of the studies discussed above ((Noesselt et al., 2010) and (Chen et al., 2011)) found sensitivity *enhancement* for even shorter audiovisual stimuli than ours. Further research could investigate this effect in a systematic way by parametrically varying perceptual performance levels, signal intensity and noise intensity. Crucially, future studies could also test for a potential role of modality specific processing latencies by introducing and parametrically varying a temporal offset between flashes and sounds.

3.4.4.2 Do longer lasting sounds affect visual sensitivity?

The pattern of results I observed can be interpreted in two major ways. One interpretation, consistent with my initial hypothesis would suppose a process lowering overall visual sensitivity in the presence of co-occurring sounds, and a second, counter-acting process of visual sensitivity enhancement for longer lasting sounds. An alternative interpretation would suppose a process lowering visual sensitivity that is exclusive to a sound of short, matching duration and would suppose no effect on visual sensitivity whatsoever for longer sounds. Although the latter interpretation appears simpler, I think the data are more in line with the first interpretation.

There are two aspects of the data that are hard to reconcile with the view that only the shortest sound duration had an effect on visual

sensitivity. The first is the shape of the curve for visual sensitivity vs. sound duration (**Figure 3-5 B**). Just as predicted, visual sensitivity gradually increased for longer durations, but fell back to baseline level for a duration purposefully chosen to fall outside the temporal window of integration (e.g. (Bolognini et al., 2005), c.f. Introduction). This drop to baseline level is expected under the hypothesis of prolonged sounds *within* the temporal window of integration enhancing visual sensitivity. But it is unexpected and hard to explain under the assumption that only sounds of matching duration had an effect on visual sensitivity. It would be interesting for future experiments to test visual sensitivity between 96 and 190 ms (for which I have no data). It would be particularly interesting to see whether visual sensitivity rises above unimodal baseline level before it drops off again.

The second aspect of the data supporting an enhancement of visual sensitivity due to prolonged sounds is the observed correlation between conditions. Across participants the maximum difference between visual baseline level and performance for prolonged visual stimuli correlated with the maximum difference between audiovisual baseline and performance for prolonged sounds. This correlation would fit with the hypothesis that prolonged sounds yield a sustain of visual perception and its underlying neural activity. I expect participants gaining more from physically prolonged visual stimulus durations to equally gain more from cross-modally induced sustain of visual representations. In contrast, there is no explanation for this correlation under the assumption that only the shortest sound duration had an effect on visual sensitivity. Taken together, I view the first of the proposed interpretations to be the more likely one for the pattern of data I

observed. But if auditory stimulus duration influences visual sensitivity, how does it do so? I will turn to this question in the general discussion of Experiments 1-3 below.

3.5 Discussion of Experiments 1-3

The main finding of Experiments 1 and 2 is that co-occurring sounds can *enhance* visual duration discrimination sensitivity if visual stimuli are accompanied by sounds of congruent durations. This extends previous findings showing that sounds of *in*congruent durations can decrease such sensitivity (Chen and Yeh, 2009; Donovan et al., 2004; Klink et al., 2011; Walker and Scott, 1981). The main finding of Experiment 3 is that visual detection sensitivity can be modulated by the duration of non-informative sounds. Detection sensitivity for a brief visual flash increased with the duration of a co-occurring sound, but only up to a limit hypothesized to reflect the width of a window of audiovisual integration.

3.5.1 Sound durations modulating visual perception

Particularly the results of Experiment 3 seem to suggest that the duration of co-occurring sounds affects the duration of visual perception itself (which implicated sustain of the underlying neural activity), rather than merely affecting a supra-modal duration specific mechanism, like an internal clock (see above; c.f. (Klink et al., 2011)). The mere presentation of auditory stimuli during a visual task can modulate visual performance (e.g. (Bolognini et al., 2005; Cappe et al., 2009; Kim et al., 2011; Leo et al., 2011;

Noesselt et al., 2010; Shams et al., 2000; Spence & Driver, 2004; Vroomen & De Gelder, 2000; Vroomen & Keetels, 2009)) as well as responses in early visual areas (Cappe et al., 2012, 2010; Lakatos et al., 2009; Murray et al., 2012; Romei et al., 2009, 2007; Wang et al., 2008). In light of these findings it seems reasonable to interpret my results as representing a modulation of the duration of visual activation corresponding to the duration of co-occurring auditory stimuli Such a mechanism would also be compatible with the results of Experiments 1 and 2. If auditory stimuli trigger a process that modulates the duration of visual perception itself, it could potentially reduce the trial-by-trial variability (i.e. error) of these durations for a given visual stimulus.

3.5.2 Potential neural mechanisms

A recent study (Romei et al., 2012) found that the presentation of a brief auditory stimulus can phase-align oscillatory activity in the alpha frequency band over occipitoparietal areas and consequently modulate perception. These findings suggest a role for alpha oscillations in determining cross-modal effects on visual cortex excitability that might apply to my results. A critical time window of ~60-100 ms (as found in Experiment 3) would correspond to one full alpha cycle and is likely to represent the temporal window for binding crossmodal information. Furthermore, it is tempting to speculate that the observed inter-individual variability in optimal duration of auditory stimuli in Experiment 3 could correspond to individual differences in oscillatory alpha frequency. Future studies should ascertain whether and to what extent the effects of auditory

stimulus duration on visual sensitivity and duration judgments are due to oscillatory phase reset.

3.5.3 Attention

An alternative (but possibly compatible) mechanism mediating the effects of longer sounds on visual sensitivity would be sound-induced attention or arousal. The accumulated stimulus energy of a longer sound will exceed the one of a shorter sound and maybe this kind of stronger signal is better suited to guide temporal attention towards the visual stimulus. Note, however that this kind of temporal uncertainty reduction has been psychophysically tested and refuted as an explanation for visual sensitivity enhancement induced by the mere presence of co-occurring sounds (Chen et al., 2011). Also, it cannot explain the results of Experiments 1 and 2. Nevertheless, it could play a role in the duration dependent effects found in Experiment 3. If temporal attention guidance is (at least part of) the mechanism behind these findings, it would point to interesting features of cross-modal integration. Its effects would be restricted to a temporal window of integration and would take place *after* the visual stimulus offset (note that the physical offset of the visual stimulus always preceded any physical differences in sound durations). The latter point underscores that any such effect of attention would be hard to distinguish from a cross-modal sustain of visual representations. The distinction between an explanation involving temporal attention and cross-modal effects might be artificial after all. Specifically, effects of cross-modal phase reset and of attention might

work hand in hand, as suggested by recent neurophysiological results (Lakatos et al., 2009).

3.5.4 Conclusion

Whatever the precise neural mechanisms behind these effects are, the results of the three experiments presented in this chapter suggest that the duration of co-occurring auditory stimuli can modulate the duration of a visual percept and thus visual perception itself. In the next chapter I will turn to another example of auditory modulation of visual perception, that is subjectively and objectively (Berger et al., 2003) quite drastic – at least for some: The sound induced flash illusion (Shams et al., 2000).

4 Are there systematic differences in brain morphology between people more or less prone to the sound induced flash illusion?

4.1 Introduction

When a single flash is accompanied by a rapid series of two or more bleeps, a perceptual 'fission' of the flash sometimes occurs and it is incorrectly perceived as multiple flashes (Shams et al., 2000). The illusion is a striking example of how sounds can modulate visual perception (and thus challenges older theories of visual dominance (Colavita, 1974; Posner et al., 1976). It is further in line with theories that propose the weighting of

sensory channels in multisensory perception depends on their relative levels of signal-to-noise (Ernst & Banks, 2002; Shams, Ma, et al., 2005). The number of events in a rapid series is easier to tell for auditory beeps than for visual flashes. This is reflected in the degree of auditory dominance over vision in the sound induced flash illusion, that thus seems statistically optimal (Shams, Ma, et al., 2005). However, perception of multisensory stimuli varies not only with stimulus properties, but also varies across observers. The same stimulus can evoke cross-modal effects reliably in some participants, but not in others. This can be seen for the sound induced flash illusion (Mishra et al., 2007), as well as for the McGurk illusion (McGurk & MacDonald, 1976). Individual differences in proneness to the McGurk illusion are correlated with the amplitude of BOLD-signal responses to cross-modal stimuli in the left superior temporal sulcus (Nath & Beauchamp, 2012; Nath, Fava, & Beauchamp, 2011). Moreover, individual proneness to the sound induced flash illusion is correlated with the degree to which sounds modulate visual event related responses (Mishra et al., 2007). However, the neural basis of this variance in proneness to audio-visual interactions is still unclear. Variability in several aspects of visual perception is correlated with differences in local brain structure (for a recent overview see (Kanai & Rees, 2011)). For instance, individual differences in the surface area of primary visual cortex are correlated with individual differences in proneness to contextual size illusions (Schwarzkopf, Song, & Rees, 2011). However, individual differences in the degree of cross-modal interactions have not previously been linked with variability in brain structure.

The objective of the experiment presented in this chapter is to test whether individual proneness to the sound induced flash illusion is correlated with differences in brain structure, using voxel based morphometry (VBM; (Ashburner & Friston, 2000)).

Experience of the 'sound induced flash illusion' is accompanied by enhanced activity in retinotopically defined primary visual cortex (V1), superior colliculus (SC), and superior temporal sulcus (STS) (Shams, Iwaki, et al., 2005; Watkins et al., 2006). Furthermore, EEG source localization (Mishra et al., 2007) and short latencies of event related magnetic field responses (ERF, (Shams, Iwaki, et al., 2005)) suggest a role of auditory cortex in the illusion. Therefore, I hypothesize that individual differences in susceptibility to the sound-induced flash illusion will be reflected in structural variation of one or more of these regions across individuals.

4.2 Methods

4.2.1 Participants

29 healthy participants from the University College London (UCL) participant pool (20 females, aged 18 to 42 years; mean: 25yrs, SD: 6yrs) took part in the study. All participants completed the behavioral study outside the scanner and underwent the anatomical MRI scan on a different day. All participants were right handed, had normal or corrected to normal vision and reported no hearing problems. Written informed consent was obtained from each participant and the study was approved by the UCL ethics committee.

4.2.2 Stimuli

The visual stimulus consisted of a uniform white disk (140 cd/m^2) that flashed for 24ms (two frames at 85Hz) on a uniform grey background (90 cd/m^2) on a CRT monitor. The disk diameter was 2° visual angle and it was placed at 5° eccentricity directly above or below a fixation cross that was displayed at the middle of the screen. The auditory stimulus consisted of a pure tone at 3.5 KHz that was played for 20ms at 65 dBA on speakers adjacent to the monitor. All stimuli were programmed and presented in MATLAB (Mathworks, Ltd.) using the Cogent Graphics (http://www.vislab.ucl.ac.uk/cogent.php) and Psychophysics Toolbox 3 extensions ((Brainard, 1997; Pelli, 1997), http://psychtoolbox.org). In each trial either one or two flashes were presented, accompanied by either no, one or two beeps, resulting in six trial types (1F0B, 2F0B, 1F1B, 2F1B, 1F2B, 2F2B, were xFxB stands for the number of flashes and beeps, respectively). The onsets of flashes and beeps were synchronous. In trials with a second flash and/or second beep the onset of the second event was time locked to 34ms after the offset of the first flash (see **Figure 4-1 a)** and **b)**).

4.2.3 Procedure

Participants sat on a chair in front of the monitor at 65cm distance. They were asked to indicate if they saw one or two flashes after each trial pressing either '1' or '2' on a numerical keypad with the index and middle finger of their right hand (in a time-window lasting 1800ms after the

stimulus presentation). Participants were advised they could ignore the beeps. Trials were presented in blocks of 102 trials with counterbalanced number of trial types in random order per block. A block lasted about five minutes and participants were encouraged to take breaks in between blocks. The position of the visual stimulus (above or below fixation) was consistent within each block and changed (counterbalanced) between blocks. Each participant completed 4-6 blocks.

To ensure that participants kept fixation throughout a block their eye movements were monitored with an eyetracker system (Cambridge Research Systems). For 21 participants eye movement data was fed into the stimulus presentation script online. For the program to present the next trial participants had to keep fixation for at least 500ms. Fixations had to be within a square region of 3°x3° around the fixation cross. For the remaining eight subjects eye-data were analyzed offline. For those participants, trials were excluded from analysis if the eyetracker did not record eye-position. Of the remaining trials (85.13%, SD: 13.54%), we included trials if participants' fixation did not deviate more than 1.5deg from the midpoint of the screen on the vertical axis (95.41%, SD: 5.01%).

4.2.4 Analysis of behavioral data

All statistical analyses of the behavioral data were performed in MATLAB (Mathworks, Ltd.) and PASW 18.0 (SPSS inc./IBM). To test for the sound induced flash illusion we compared the proportion of correct answers between conditions with a repeated measures ANOVA and post-hoc t-tests. To determine proneness to the illusion a 'fission score' (FiS) was calculated

for each subject. It was defined as one minus the proportion of correct answers in the illusion trials (pCorr(1F2B)) and corrected for any response bias to report two flashes independent of the number of beeps:

$$FiS = (1 - pCorr(1F2B)) * (pCorr(1F0B) / pCorr(2F0B))$$

Additionally, the fission score was calculated separately for the two stimulus positions (above or below fixation cross). The mean, range and variance for FiS and the simple proportion of correct trials were determined. Furthermore the correlation between FiS for trials in which the disk flashed above and below the fixation cross, respectively, was calculated.

4.2.5 MRI data acquisition and pre-processing

T1 anatomical images of the brain were obtained with a 1.5 Tesla Siemens Sonata MRI scanner (Siemens Medical). High-resolution anatomical images were acquired using a T1-weighted three-dimensional modified driven equilibrium Fourier transform sequence (MDEFT; repetition time = 12.24 ms; echo time = 3.56 ms; field of view = 256 × 256 mm; voxel size = 1 × 1 × 1 mm).

T1-weighted MR images were first segmented for grey matter (GM) and white matter (WM) using the 'New Segment' segmentation tools in Statistical Parametric Mapping 8 (SPM8, http://www.fil.ion.ucl.ac.uk/spm). Subsequently, we performed diffeomorphic anatomical registration through exponentiated lie algebra (DARTEL) in SPM8 for intersubject registration of the GM and WM images (Ashburner, 2007). To ensure that regional grey

matter volume was maintained after the registration, the registered images were modulated by the Jacobian determinant of the flow fields computed by DARTEL. The registered images were smoothed with a Gaussian kernel of 8 mm full-width at half-maximum and transformed to Montreal Neurological Institute (MNI) stereotactic space using affine and nonlinear spatial normalization implemented in SPM8.

4.2.6 Voxel based morphometry: Statistical analysis

To test for correlations between grey matter volume and illusion strength, multiple regression analyses were performed on the smoothed grey matter images. Fission scores were entered as vectors of interest into the design matrix, while total grey matter volume, age and sex were included as regressors of no interest in the model to control for any differences in these variables.

To incorporate our *a priori* hypotheses concerning the brain structures we predicted to be involved, region of interest masks were created using the SPM anatomy toolbox (http://www.fz-juelich.de/inm/inm-1/spm_anatomy_toolbox) and MarsBaR (http://marsbar.sourceforge.net/). The first three regions of interest were derived from the illusion-specific significant activations reported in (Watkins et al., 2006). This study reported significant activation of retinotopically defined V1, of right posterior STS and the superior colliculus. Since we did not have retinotopic data for our VBM subjects and the size of V1 shows relatively large inter-individual variation (Dougherty et al., 2003; Schwarzkopf et al., 2011), we used histological maximum probability maps

to combine the BA17 and BA18 regions into one mask (Eickhoff, Heim, Zilles, & Amunts, 2006). Right posterior STS and SC were incorporated via a 10mm and 4mm radius sphere, respectively, centered on the stereotactic coordinates of the peak voxels reported in (Watkins et al., 2006): [54,-54,30] and [2, -30, 0]. Primary auditory cortex was added as a region of interest following (Mishra et al., 2007) and as defined by histological maximum probability maps (Eickhoff et al., 2006).

Grey matter volumes were logit transformed (Ashburner & Friston, 2000) and the average grey matter volume within the regions of interest was derived with MarsBaR and correlated with fission scores (controlling for total grey matter volume, age and sex). We used Bonferroni correction to adjust statistical thresholds for multiple ROIs tested. Outside the ROIs, an additional exploratory whole brain analysis was performed, using a threshold of $p < 0.05$, corrected for multiple comparisons using the family-wise error rate (FWE).

4.3 Results

4.3.1 Behavioral results

The sound induced flash illusion was replicated (Shams et al., 2000; Watkins et al., 2006). Participants on average answered correctly on 38% (SD: 28%) of 1F2B trials indicating that they perceived the illusion on average on 62% of trials. The inter-individual range was 2-100%. A repeated measures ANOVA indicated a significant difference between conditions ($F_{(2.50)}$=45.17, p<.001, Greenhouse-Geisser corrected; **Figure 4-1**

c); **Table 4-1**). The proportion of correct answers was significantly lower in the 1F2B condition, as compared with all other conditions and indicated by post-hoc paired *t*-tests. 95% confidence intervals of the differences indicated this was a strong and robust effect (Table 2). Whether the disc appeared in the upper or lower visual field did not alter the strength of the sound induced flash illusion (fission score below: 0.64 (SD: 0.29), fission score above: 0.62 (SD: 0.36); paired t-test: $t_{(28)}$=0.64, p=.53 *n.s.*). The probability of reporting the illusion was highly correlated within individuals and across trials in which the disk flashed above vs. below the fixation cross (r=.84, p<10^-7). The distribution of fission scores across participants is given in **Figure 4-1 d)**.

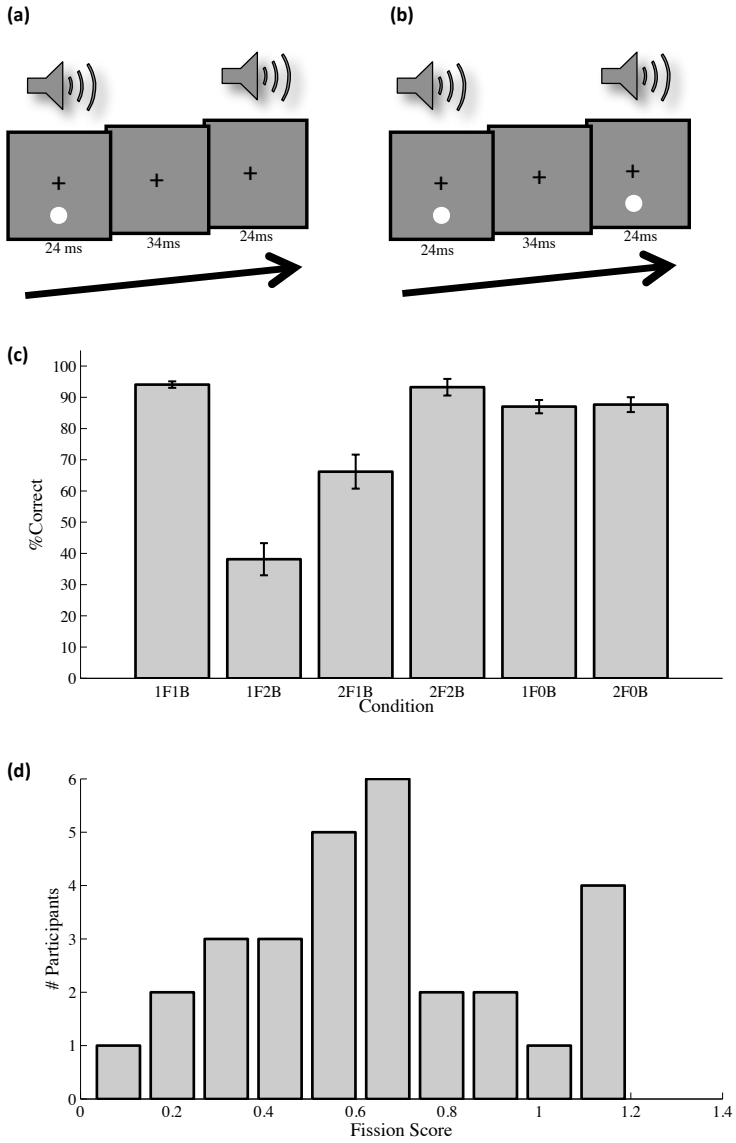

Figure 4-1 Stimulus sequences and behavioral results. The visual stimulus consisted of a visual disk (with a radius of one degree visual angle) that flashed once or twice at 5 degrees eccentricity below or above the fixation cross, which was placed at the middle of the screen. The flash or flashes were combined with zero, one or two beeps (3.5 KHz at 65 dBA). Panel **(a)** illustrates the critical trial type for the sound induced flash illusion: One visual flash is accompanied by two beeps (1F2B). Panel **(b)** illustrates a trial in which two flashes were accompanied by two beeps (2F2B). Panel **(c)** shows behavioral results by trial type. In each trial participants (n=29) indicated whether they saw one or two flashes. Bars represent the percentage of correct answers by trial type, averaged across participants. xFxB refers to number of flashes and beeps, respectively (from left to right: trials with one flash and one beep; one flash and two beeps; two flashes and one beep; two flashes and two beeps; one flash only; two flashes only). Error bars represent the standard error of the mean. Note that the second bar represents the critical trial type (one flash, two beeps). The low overall proportion of correct answers for this trial type (38%) indicates that participants perceived the illusory second flash in the remaining trials of this condition (cf. **Table 4-2**). Panel **(d)** shows the distribution of fission scores across the sample.

Table 4-1 Descriptive statistics for behavioural data. Cells contain information regarding the proportion of correct answers by trial type (xFxB refers to number of flashes and bleeps, respectively; FiS: Fission Score (see methods); below/above: position of flashing disk relative to fixation cross, see methods; SD: standard deviation).

	Min	Max	Range	Mean	SD
1F1B	0.79	1.00	0.21	0.94	0.06
1F2B	0.00	0.98	0.98	0.38	0.28
2F1B	0.04	0.98	0.94	0.66	0.29
2F2B	0.32	1.00	0.68	0.93	0.14
1F0B	0.57	1.00	0.43	0.87	0.11
2F0B	0.49	1.00	0.51	0.88	0.13
FiS	0.02	1.20	1.17	0.63	0.30
FiS below	0.00	1.25	1.25	0.64	0.29
FiS above	0.04	1.39	1.35	0.62	0.36

Table 4-2 Group analysis of illusion effect. *t*-statistics with corresponding sample standard deviation, *p* values and 95% confidence intervals for paired *t*-tests contrasting the 1F2B condition with all other conditions. Note that the proportion of correct answers was significantly lower compared with all other conditions (cf. Fig. 1 C). 95% confidence intervals of the difference are given as [lower boundary upper boundary].

Contrast	$t_{(28)}$	*sd*	95% CI	*p*
1F2B vs. 1F1B	-11.24	0.27	[-0.66 -0.46]	<10^-11
1F2B vs. 2F1B	-4.86	0.31	[-0.40 -0.16]	<10^-4
1F2B vs. 2F2B	-9.11	0.33	[-0.67 -0.42]	<10^-9
1F2B vs. 1F0B	-10.01	0.26	[-0.59 -0.39]	<10^-10
1F2B vs. 2F0B	-10.07	0.26	[-0.60 -0.39]	<10^-10

4.3.2 MRI results

The VBM analysis revealed a strong and statistically significant negative correlation between fission scores and local grey matter volume in the BA17&18 region of interest (controlled for global grey matter volume, age and sex): r=-.55, $t_{(24)}$=-3.27, p=.01 (two-tailed and Bonferroni corrected for multiple ROIs; cf. **Figure 4-2**). Note that this correlation remained qualitatively unchanged and statistically significant when not controlling for age and gender (r=-.54, $t_{(26)}$=-3.30, p=.003) and when using raw behavioral scores (1-pCorr(1F2B)) instead of FiS (r=-.47, $t_{(26)}$=-2.74, p=.01). No significant correlation between local grey matter volume and fission scores was found for the primary auditory cortex, posterior STS and superior colliculus regions of interest (cf. **Figure 4-3**.). An additional exploratory whole brain analysis yielded no further significant findings at a threshold of P<0.05, corrected for multiple comparisons (family wise error correction).

To further test whether the correlation between proneness to the fission illusion and grey matter volume in the BA17&18 region of interest was driven by the BA17 or BA18 region (or both), we correlated grey matter volume in each region separately with fission scores (again controlled for global grey matter volume, age and sex). The grey matter volume of both regions was significantly negatively correlated with fission scores; the relationship was slightly stronger for the BA18 mask (r=-.60, $t_{(24)}$=-3.68, p=.001, two-tailed) than for the BA17 mask (r=-.47, $t_{(24)}$=-2.64, p=.01, two-tailed), but this difference was not significant ($Z_{(}$=0.97, p=.33, *n.s.*).

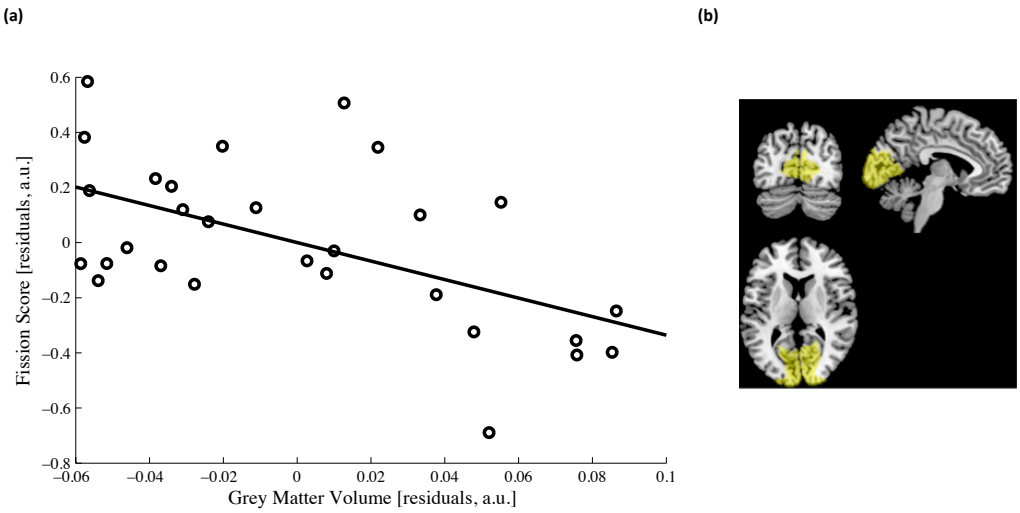

Figure 4-2 Correlation between proneness to the sound induced flash illusion and grey matter volume in early visual cortex. Each circle in (**a**) represents the BA17&18 grey matter volume and Fission Score of one participant (see methods for details of Fission score). The plot shows residuals after controlling for total grey matter volume, age and sex. Note that not controlling for age and sex, and using raw behavioural scores rather than Fission Scores left the correlation qualitatively unchanged and statistically significant. The image to the right (**b**) shows the corresponding BA17&18 region of interest projected on slices from the coronal, sagittal and axial planes of a canonical T1 weighted structural image ('collin27', (Holmes et al., 1998)). The image is in Montreal Neurological Institute (MNI) stereotactic space. The mask was derived using histological maximum probability maps to combine the BA17 and BA18 regions into one mask (Eickhoff et al., 2006).

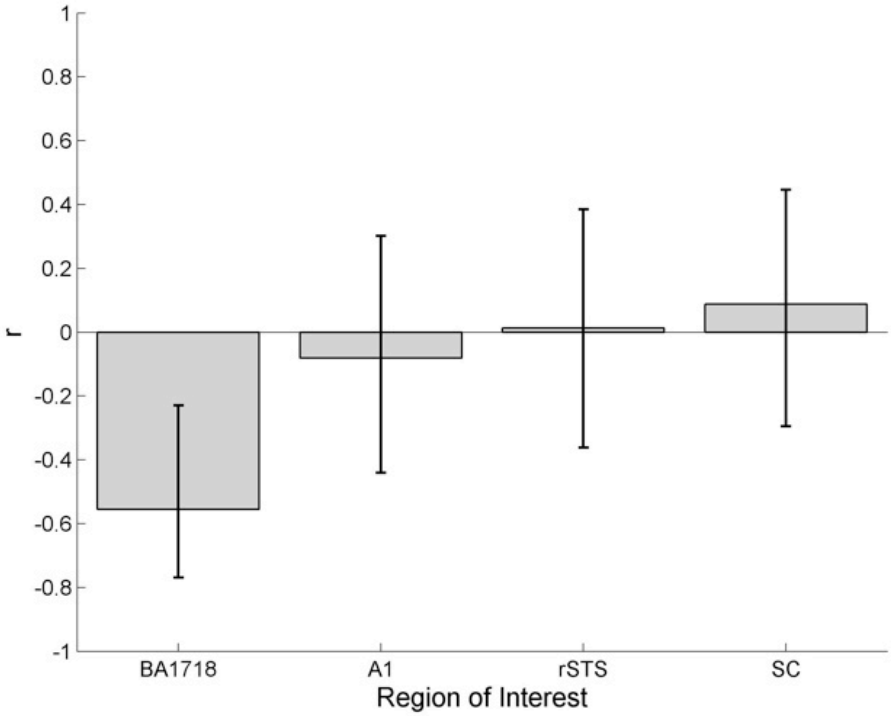

Figure 4-3 Correlation between proneness to the sound induced flash illusion and grey matter volume in regions of interest. Bars indicate Pearson correlation between proneness to the sound-induced flash illusion and grey matter volume in regions of interest (controlled for total grey matter volume, age and sex). Values on the y-axis reflect partial correlation coefficients. Error bars indicate 95% CIs of r, estimated using Fisher's z-transformation. Regions of interest were BA17/18: early visual cortex (BA17 and 18 regions derived using Juelich Histological Atlas, see text); A1: primary auditory cortex (derived using Juelich Histological Atlas, see text); rSTS, right superior temporal sulcus (10 mm sphere centred on peak voxel reported by (Watkins et al., 2006); see text); SC, superior colliculus (4 mm sphere centred on peak voxel reported by (Watkins et al., 2006); see text).

4.4 Discussion

I found reliable inter-individual differences in proneness to the sound induced flash illusion. While some participants experienced the illusion hardly ever, others experienced it on almost every trial. Moreover,

individual proneness to the illusion proved reliable across blocks with different flash locations, suggesting it to be a stable, trait-like feature.

These individual differences in proneness to the illusion were correlated with local grey matter volume in early visual cortex (cf. **Figure 4-2**). Participants with low grey matter volume in the BA17/18 region experienced the illusion significantly more often. Given I regressed out global grey matter volume, this points to a systematic relationship between individual proneness to the illusion and the relative amount of total grey matter dedicated to early visual cortex.

4.4.1 Potential neural mechanisms

4.4.1.1 Visual cortex activity

Higher proneness to the sound induced flash illusion is associated with greater multisensory modulation of visual ERPs (Mishra et al., 2007). In that earlier study proneness to the illusion correlated not only with the magnitude of illusion associated ERPs but also with supra-additive multisensory ERPs in trials with two beeps and two flashes (that did not induce any illusion). This result suggests increased proneness to the sound induced flash illusion is associated with a more general enhancement of audiovisual integration. BOLD responses in V1 were found to be enhanced in an illusion specific way (Watkins et al., 2006) . That is, BOLD responses in V1 were only enhanced by a second beep in trials in which it successfully induced the illusion. Taken together, these previous studies demonstrate that proneness to the sound induced flash illusion correlates with stronger

and more frequent multisensory modulations of early visual cortex activity. In light of these previous findings and the results of the current experiment, it seems likely that individuals with smaller visual cortices will exhibit stronger multisensory modulation of visual responses. This hypothesis could be tested directly, for instance by comparing individual differences in the surface area of V1/V2 and modulations of BOLD responses in these areas related to the sound induced flash illusion (c.f. (Watkins et al., 2006)).

4.4.1.2 Visual cortex anatomy

One possible explanation for the current observation of greater multisensory perceptual effects for observers with anatomically smaller visual cortices may be incomplete scaling of multisensory connections with early visual cortex. Because visual cortex volume was effectively normalized for total grey matter volume in the current study, a smaller value implies a smaller proportion of grey matter dedicated to early visual processing. This might imply a greater ratio of multisensory grey matter (e.g. in superior temporal sulcus) vs. visual grey matter and thus result in a greater number of multisensory synapses per visual neuron. This, in turn, would explain the higher likelihood for auditory modulation of visual perception, as indicated by proneness to the sound induced flash illusion. However, this hypothesis should be taken with care and needs further testing. A post-hoc test on the present data could not confirm a significant correlation between fission score on the one hand and the ratio of grey matter volume in the BA1718 mask vs. in the spherical right posterior STS mask on the other (r=-.18, $t_{(25)}$=-0.91, p=.37, *n.s.*).

Variability in V1 surface area is negatively correlated with proneness to contextual visual size illusions (Schwarzkopf et al., 2011). This result is of interest in the context of my study, because it parallels the relationship between small visual cortex and high illusion proneness I found. This suggests that contextual influences may be generally increased in small visual cortices – both, within and across modalities. (Schwarzkopf et al., 2011) interpreted their finding as pointing to a greater number of lateral connections from distant visual field representations within smaller visual cortices. One might speculate that my results might point to a similar neuroanatomical phenomenon, albeit across different areas of the brain rather than within one area. Increased contextual influence on visual processing might be due to a higher degree of neural connectedness in small visual cortices –within and across areas as well as within and across modalities.

4.4.2 Reliability-based weighting of sensory channels

An alternate hypothesis is that the weighting of sensory channels is tuned to the availability of neural resources. Our finding suggests that the relative amount of neural resources dedicated to the visual modality *in an individual brain* influences the weight placed on this sensory channel. Such a mechanism would be complementary to weighting mechanisms tuned to relative levels of input noise (Battaglia, Jacobs, & Aslin, 2003; Beauchamp, Pasalar, & Ro, 2010; Ernst & Banks, 2002; Morgan, Deangelis, & Angelaki, 2008) as confirmed for the sound induced flash illusion (Shams, Ma, et al., 2005). If the brain weights sensory channels according to their relative

levels of *effective* noise, this will reflect more than input noise. It will also take into account the relative levels of *intrinsic* noise of sensory channels. One (well-studied) aspect of such intrinsic noise is the general suitability of a sensory channel for the stimulus dimension at hand. This is reflected in general tendencies across participants, like visual dominance for spatial judgments and auditory dominance for temporal judgments (e.g. (Alais & Burr, 2004); (Romei et al., 2011)). I.e. the signal to noise ratio for spatial stimulus aspects is generally higher in the visual than in the auditory channel, whereas the opposite is true for temporal stimulus aspects.

My results suggest an additional, more subtle aspect of intrinsic noise: it may vary between subjects according to the amount of grey matter dedicated to the specific sensory channel. Such a hypothesis would be in line with previous results, showing that the surface area of primary visual cortex in healthy humans is correlated with the cortical magnification factor at eccentricities comparable to that of our visual stimulus (Harvey & Dumoulin, 2011). That is, subjects with smaller visual cortices have a visual representation that exhibits coarser spatial tuning. Further, this relationship between visual cortex area and acuity exists on the behavioral level as well. Cortical magnification within V1 is correlated with Vernier and grating acuity thresholds – across observers and eccentricities (Duncan & Boynton, 2003). Taken together, previous results have shown that early visual cortex size correlates with the spatial resolution of visual representations. It is tempting to speculate that this points to a general correlation between early visual cortex size and the signal to noise ratio of visual representations (including temporal resolution). Under this assumption the current finding

suggests an intriguing possibility: The weighting of the visual channel in multisensory integration might be tuned to the amount of grey matter dedicated to early visual cortex.

4.4.3 Attention

Finally, my results might be linked to individual differences in attention mechanisms (Kanai, Dong, Bahrami, & Rees, 2011; Kanai & Rees, 2011). Recent findings link the strength of the illusion to several such mechanisms. Specifically, top-down modality-specific attention shifts can suppress processing in the distractor modality and thereby attenuate the (visuotactile version of the) illusion (Werkhoven, van Erp, & Philippi, 2009). Spatial attention directed away from the audiovisual stimuli diminishes early occipitotemporal components of the illusion specific ERP components (which have been shown to be increased for participants experiencing the illusion more often (Mishra, Martínez, & Hillyard, 2010)). Disruption of the angular gyrus with TMS results in less frequent perception of the illusion, which has been attributed to attenuated effects of bottom-up attention evoked by the sounds (Kamke, Vieth, Cottrell, & Mattingley, 2012). Consequently, participants who are more prone to the sound induced flash illusion could be more susceptible to auditory attentional capture, they could allocate more attention to the spatial position of the audiovisual stimuli, or they could be less able to suppress the auditory modality via top-down attention. It is also interesting to speculate whether both, effects of cross-modal attention and early visual cortex grey matter volume may be

linked to effects of large network oscillatory phase-reset (Lakatos et al., 2009; Romei et al., 2012).

4.4.4 Future experiments

Future experiments could shed more light on the mechanisms behind the present finding. A putative relationship between early visual cortex size and its structural connectivity with multisensory areas can be tested using probabilistic tractography (Beer, Plank, & Greenlee, 2011). Functional definitions of early visual cortex (using retinotopic mapping, (Sereno et al., 1995)) would allow for accurate delineation of visual areas on the cortical surface of the individual brain. The surface-based analysis necessary for this would also allow to dissociate visual cortex area and thickness and thus to test their respective association with proneness to the sound induced flash illusion. To test the role of subtle differences in visual processing across participants, future studies could employ sensitive visual tests, such as Vernier acuity or a version of our purely visual trials, modified to enhance inter-individual variance in this condition (such as adding noise masks). A potential link between proneness to the illusion and differences in cross-modal attention could be tested behaviorally and followed up by tests on the individual propensity for cross-modal oscillatory phase reset (Lakatos et al., 2009; Romei et al., 2012).

4.4.5 Conclusion

In summary, I found a strong, negative correlation between early visual cortex grey matter volume and proneness to the sound induced flash illusion. I proposed a neuroanatomical and functional explanation for this finding and ways to test these explanations in further experiments.

In the next chapter I will move away from the use of artificial flash and beep-type stimuli to probe potential auditory modulations of visual cortex activity when viewing ecologically valid video-type stimuli.

5 Does auditory input modulate visual stimulus encoding in V1-3?

5.1 Introduction

Perception of the environment requires integration of sensory information across the senses, but how our brains combine information from different sensory streams is still poorly understood. The classic view of multisensory integration holds that cortical sensory processing falls in two parts: The sensory input is first processed in strictly unimodal, primary sensory areas. Multisensory processing then takes place in and is restricted to dedicated convergence areas (Mesulam, 1998). However, this view has changed in the past decade due to new anatomical and functional evidence for multisensory interactions at the level of primary sensory areas. While neurons in these areas typically cannot be driven by other modalities their activity might be modulated by stimulation of another modality (c.f. general

introduction for more details; see (Driver & Noesselt, 2008) and (Klemen & Chambers, 2012) for an overview).

The understanding of how our brains combine information across different senses is especially poor when it comes to ecologically valid, dynamic stimuli. Two recent monkey studies used such naturalistic video-type stimuli to investigate how processing in the primary auditory cortex is modulated by accompanying visual stimulation (Kayser et al., 2010) and how the activity of visual neurons in the superior temporal sulcus (STS) is affected by auditory co-stimulation (Dahl et al., 2010). The main finding of both studies is that bimodal stimulation affected the inter-trial reliability of the spike trains of the 'unimodal' neurons and thus the amount of stimulus-information that could be retrieved from these spike patterns. (Kayser et al., 2010) found that congruent visual stimulation (matching videos) significantly enhanced stimulus information carried by spike trains from auditory cortex. This was true in comparison with unimodal auditory stimulation and incongruent visual co-stimulation (mismatching videos). (Dahl et al., 2010) found that incongruent sounds significantly worsened stimulus decoding from visual neurons in STS, compared to unimodal visual and audiovisual congruent stimulation.

In the experiment presented in this chapter I sought to test whether multisensory modulation of stimulus encoding extends to humans and early retinotopic visual cortices. Because I could not use single cell recording with healthy human participants I had to ask a version of this question that is slightly different from the operationalization in the monkey studies discussed above. These studies probed the reliability and amount of

stimulus information carried by *temporal* patterns of activity of single neurons. Instead, the experiment presented here probed the reliability and discriminability of stimulus-evoked *spatial* patterns of activity in V1-3 using multivoxel pattern analysis (MVPA, c.f. Chapter 2.5.4).

I presented participants with naturalistic audiovisual stimuli in four different conditions: audio only (A), visual only (V), audiovisual congruent (AV congruent) and audio-visual incongruent (AV incongruent). I then used MVPA to decode stimulus identities based on spatial patterns of BOLD signals evoked in V1-3 (as identified by retinotopic mapping, (Sereno et al., 1995), c.f. General Methods). Separate multivariate classifiers were trained and tested for each of the four conditions and for each ROI. This allowed me to compare decoding accuracies between conditions, thus obtaining an index of pattern discriminability for each condition.

5.2 Methods

5.2.1 Participants

15 healthy participants from the University College London (UCL) participant pool took part in the experiment (mean age, 26 yrs, SD, 4 yrs; 7 females; 1 left handed). All participants had normal or corrected to normal vision and reported no hearing problems. Written informed consent was obtained from each participant and the study was approved by the UCL ethics committee. Participants were paid 10 GBP per hour for taking part in the experiment, which lasted up to 2.5 hours.

5.2.2 Stimuli

Four video clips were used as audio-visual stimuli, each lasting 3 s. Two clips showed natural scenes containing animals (a croaking frog and a crowing rooster). These two clips were downloaded from http://www.youtube.com and edited. The two remaining clips showed the clothed torso of the author while turning a key in a lock or ripping a paper apart. All clips were similar with regard to luminance and loudness and were projected onto a screen at the end of the scanner bore. Participants viewed the clips via a mirror mounted at the head coil of the scanner at a viewing distance of ~72 cm. Video clips were presented at a resolution of 640x360 pixels and subtended ~18 by 10 degrees visual angle when viewed by participants in the scanner. During the experiment participants were asked to fixate a white dot projected on top of the videos at the center of the screen (radius ~0.1 degree visual angle). In each trial the dot turned blue once, twice or three times and participants were asked to count and indicate the number of color changes via a button box in a 2s inter stimulus interval.

Audio tracks accompanying the video clips were presented via MRI compatible in-ear headphones (http://www.etymotic.com). Loudness was adjusted individually before the start of the experiment, aiming for a level that was comfortable for participants but still enabled them to easily tell apart sound clips in the presence of scanner noise.

All stimuli were programmed and presented in MATLAB (Mathworks, Ltd.) using the Cogent Graphics (http://www.vislab.ucl.ac.uk/cogent.php) and Psychophysics Toolbox 3 extensions (Brainard, 1997; Pelli, 1997; http://psychtoolbox.org).

5.2.3 Procedure

Each participant completed 17-24 runs of scanning in the main experiment, each run lasting just under 2 minutes. During the runs participants were presented with audio and/or visual clips and completed an incidental, superimposed fixation task (cf. above). During each run each of the 4 stimuli was presented once for each experimental condition (i.e. four times), amounting to 16 stimulus trials per run (cf. **Figure 5-1**). Participants were either presented with videos only (V), sounds only (A), matching videos and sounds (AV congruent condition) or mismatching videos and sounds (AV incongruent condition). For audio-visually incongruent trials the sound files were swapped between fixed pairs of videos (rooster crowing and paper ripping; frog croaking and keys turning). Each 3 s clip was followed by a 2 s inter-stimulus interval during which participants were asked to indicate via a button box how many times the fixation dot changed its color. In addition to the 16 stimulus trials there were 4 blank trials in each run that served as a baseline measure. During these trials participants completed the fixation task in the absence of audio-visual clips. The order of the 20 trials was randomized for each run, as was the number of fixation dot color changes in each trial (1-3).

5.2.4 Retinotopic mapping

To delineate the borders of visual areas V1-3 on an individual basis, each participant underwent an additional fMRI run viewing stimuli for

phase encoded retinotopic mapping (Sereno et al., 1995). Stimuli for this run consisted of a wedge rotating clock-wise and an expanding ring. Both stimuli moved in discrete steps, synchronized with the acquisition of fMRI volumes, but with different frequencies (wedge: 12 cycles, 20 steps per cycle; ring: 20 cycles, 12 steps per cycle). They were centered on a fixation dot of ~0.25 degrees diameter and spanned up to 8 degrees of eccentricity. It is generally difficult to distinguish retinotopic maps inside the foveal confluence because the borders between regions are difficult to resolve at conventional voxel sizes. Moreover, the presence of a stable fixation dot precludes any systematic variation in the BOLD signal related to the mapping stimulus. Note that the size of the fixation dot for the mapping stimuli was slightly larger than the size of the fixation dot for the audiovisual stimuli (~0.25 vs. ~0.2 degrees diameter). Thus the region of interest analyses did not include the foveal representations. Ring and wedge were presented on a grey background and served as apertures revealing a dynamic high contrast stimulus. Participants were asked to fixate at all times and count brief color changes of the fixation dot from blue to purple. These color change events lasted 200 ms and could occur at every non-consecutive 200 ms window of the run with a probability of 5 %.

5.2.5 Image acquisition and pre-processing

All functional and structural scans were obtained with a Tim Trio 3T scanner (Siemens Medical Systems, Erlangen, Germany), using a 12-channel head coil. Functional images for the main experiment were acquired with a gradient echo planar imaging (EPI) sequence (3 mm isotropic resolution,

matrix size 64 x 64, 40 transverse slices per volume, acquired in ascending order (whole head coverage); slice acquisition time 68 ms, TE 30 ms, TR 2.72 s). I obtained 42 volumes per run of the main experiment (including three dummy volumes at the beginning of each run and two at the end), resulting in a run duration of 114.24 s. Functional images for retinotopic mapping were acquired in one run of 247 volumes with an EPI sequence (including five dummy volumes at the beginning and two at the end of the run; 2.3 mm isotropic resolution, matrix size 96 x 96, 36 transverse slices per volume, acquired in interleaved order (centered on the occipital cortex); slice acquisition time 85 ms, TE 36 ms, TR 3.06 s per volume). In between the main experiment and the retinotopic mapping run I acquired fieldmaps to correct for geometric distortions in the functional images caused by heterogeneities in the B0 magnetic field (double-echo FLASH sequence with a short TE of 10 ms and a long sequence of 12.46 ms, 3x3x2 mm, 1 mm gap). Finally, I acquired a T1-weighted structural image of each participant using an MDEFT sequence ((Deichmann, Schwarzbauer, & Turner, 2004); 1 mm isotropic resolution, matrix size 256 x 240, 176 sagittal slices, TE 2.48 ms, TR 7.92 ms, TI 910 ms).

All image files were converted to NIfTI format and pre-processed using SPM 8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). The dummy volumes for each run were discarded to allow for the T1 signal to reach steady state. The remaining functional images of the main experiment and the retinotopic mapping session were independently mean bias corrected, realigned and unwarped (using voxel displacement maps generated from the fieldmaps). Finally the functional images were co-

registered with the respective anatomical MDEFT scan for each participant and smoothed with a 5 mm Gaussian kernel.

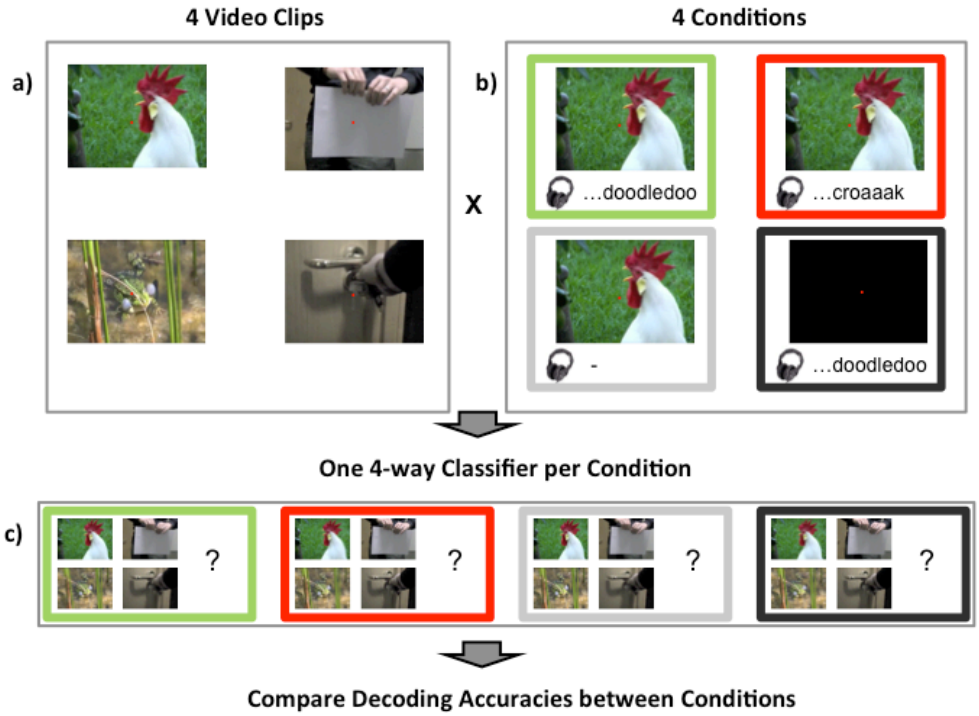

Figure 5-1 Design. a) Four audiovisual clips used as stimuli, each lasting 3s. Participants counted color changes of the fixation dot in each trial. **b)** Each of the clips was presented multiple times in four conditions (illustrated here for one example clip): audiovisual congruent (AV congruent) in green, audiovisual incongruent (AV incongruent) in red, visual only (V) in light grey and audio only (A) in dark grey. **c)** Separate multivariate classifiers were trained to decode which of the four stimuli was presented for each condition.

5.2.6 Data analysis

5.2.6.1 Multivoxel pattern analysis

I specified separate general linear models for each run and each participant. Each general linear model contained regressors for each of the 16 trial types plus one regressor for the blank trials (boxcar regressors convolved with a canonical hemodynamic response function). Additional regressors of no interest were modeled for response intervals and for the six motion parameters estimated during re-alignment. The general linear models for each run and each participant were estimated and contrast images for each of the 16 trials per run were calculated, yielding one contrast image per trial of the experiment. This resulted in separate contrast images and t-maps for each trial type of the experiment for each participant. These t-maps were masked with the retinotopic regions of interest (see below) and the resulting patterns were vectorised (i.e. collapsed into a single column of data with entries corresponding to voxels in the original data space). For the decoding and correlation analyses the resulting patterns were mean corrected across stimuli within each condition. Note that this did not affect classification performance – the distribution of patterns in feature space was preserved, but now centered on zero. This allowed me to ensure that any common intercept of patterns across stimuli was disregarded for the similarity and reliability correlation analyses (see below). Beta maps for univariate analyses were not mean corrected. The aim of the decoding analysis was to decode *stimulus identity* from activation patterns in visual areas (i.e. which of the four videos was presented in a given trial) and to compare the accuracies of decoders *across conditions* (i.e.

did stimulus decoding accuracy vary depending on audiovisual condition, cf. **Figure 5-1**). Stimulus decoding was performed using custom code using the linear support vector machine (lSVM) implemented in the Bioinformatics toolbox for MATLAB (version R2010b, http://www.mathworks.com). Data from each condition were used for training and testing of separate classifiers to get condition-specific decoding accuracies. For each condition a four-way classifier was built, to decode which of the four stimuli was presented from a given activation pattern. The four-way classifier consisted of six lSVMs to test all possible pair-wise comparisons between the four stimuli. It then assigned one of the stimulus labels based on a one-against-one voting procedure (Hsu & Lin, 2002). The four-way classifier was trained and tested for accuracy in a jackknife procedure. In each iteration, the (condition-specific) data from all runs but one served as training data and the (condition-specific) data from the remaining run was used to test the prediction accuracy of the lSVM. Accuracies were stored and averaged across iterations at the end of this procedure, and the whole procedure was applied to each retinotopic ROI (V1-3) independently, yielding a four-way classification accuracy for each condition and ROI. Statistical analysis of the resulting accuracies was done in MATLAB and PASW 18.0 (SPSS inc./IBM). Accuracies were compared against chance level by subtracting .25 and using one sample t-tests. Accuracies were compared between conditions using ANOVAs and paired t-tests.

Potential differences in decoding accuracy between conditions could stem from two different sources. They could be due to changes in pattern reliability across trials or to changes in pattern similarity between patterns

evoked by different stimuli or both. We employed additional analyses to differentiate between those options. Changes in pattern reliability were tested by averaging the patterns for a given stimulus across trials separately from odd and even runs and computing the Pearson correlation coefficient for the two resulting mean patterns (in a ROI- and condition-specific manner). The resulting correlation coefficients were Fisher z-transformed, averaged for each condition and then compared across conditions using ANOVAs and paired t-tests. Changes in pattern similarity were tested by averaging the patterns for a given stimulus across all trials and computing correlations between these mean patterns for different stimuli (again, in a ROI- and condition-specific manner). The resulting Pearson correlation coefficients were compared as described above.

5.2.6.2 Searchlight analysis

To test whether and where stimulus information was modulated by audiovisual context outside retinotopic cortices, I set up an additional, exploratory searchlight analysis (Kriegeskorte et al., 2006). For this analysis, activation patterns were derived from the same (trial-specific) t-maps that were used for the ROI analysis described above. The searchlight consisted of a sphere with a radius of 4 voxels that was centered on each grey matter voxel of each participant's brain in turn. During each iteration, the searchlight was used as a mask and the patterns of activation within this mask were read out for each trial. Then the same 4-way classification procedure used for the ROI analysis was applied to those patterns (cf. above). The resulting (condition specific) classification accuracies were

projected back onto the seed voxel. Repeating this procedure for every grey matter voxel, I thus derived four accuracy maps for each participant (one per condition). To test for significant accuracy differences between conditions I subtracted the respective accuracy maps from each other. Specifically, I contrasted the audiovisual congruent condition with the muted condition and with the incongruent condition and the muted condition with the audio-visual incongruent condition. The resulting accuracy contrast maps were normalized to MNI space (http://www.loni.ucla.edu/ICBM/) and tested for whole brain family-wise error (FWE) corrected significance at cluster level in SPM 8 (cluster forming threshold $p < .001$ uncorrected). Significant clusters were identified anatomically using the Juelich Histological Atlas implemented in the SPM Anatomy Toolbox (v. 1.8, http://www.fz-juelich.de/inm/inm-1/DE/Forschung/_docs/SPMAnatomyToolbox/SPMAnatomyToolbox_node.html).

5.2.6.3 Univariate analysis

To test whether audio-visual context had any influence on the overall signal amplitude in our ROIs I employed an additional mass-univariate analysis. For this analysis I averaged the condition specific beta weights of voxels within our ROIs across stimuli and trials for each participant. We then compared the mean beta values between conditions for each ROI using ANOVAs and paired t-tests.

I additionally tested whether a different, more traditional approach to univariate analyses would have yielded any differences between

conditions. To test this, I concatenated all runs of a given participant in one design matrix in SPM8. This allowed me to build contrasts between conditions on the first level, utilizing all trials of the respective conditions. These first level contrasts were then normalized to MNI space and tested for whole brain FWE corrected significance at cluster level in SPM8 (cluster forming threshold $p < .001$ uncorrected).

5.2.6.4 Retinotopic mapping

Retinotopic ROIs were identified using standard phase-encoded retinotopic mapping procedures (Sereno et al., 1995). I extracted and normalized the time series for each voxel and applied a fast Fourier transformation to it. Visually responsive voxels were identified by peaks in their power spectra that corresponded to our stimulus frequencies. The preferred polar angle and eccentricity of each voxel was then identified as the phase lag of the signal at the corresponding stimulus frequency (wedge and ring, respectively). The phase lags for each voxel were stored in a 'polar' and an 'eccentricity' volume and then projected onto the reconstructed, inflated cortical surface (surface based analysis was performed using FreeSurfer: http://surfer.nmr.mgh.harvard.edu). The resulting maps allowed me to identify meridian polar angle reversals and thus to delineate the borders of visual areas V1-3 on the cortical surface. These labels were then exported as three-dimensional masks into NIfTI space and served as ROIs.

5.3 Results

5.3.1 Behavioral data

Participants performed well on the fixation task for all four stimulus categories and the baseline category. Performance did not differ significantly between conditions (note that the task was independent of stimulus category; 95+/-1%, 96+/-1%, 96+/-1%, 97+/-1%, 97+/-1 % correct for the AV congruent, AV incongruent, V, A and baseline category, respectively (mean +/- standard error of the mean); $F_{(2.49, 34.85)} = 1.59$, $p = .22$, *n.s.*, Greenhouse-Geisser corrected for non-sphericity).

5.3.2 Multivariate fMRI results

5.3.2.1 Multivariate ROI results

Visual stimulus identities could be decoded significantly above chance level (0.25) from V1-3 (ROIs were combined across hemispheres; all $p < 10^{-5}$, cf. **Figure 5-2 a)**). When no visual stimulus was presented (A condition) decoding performance was at chance level (all $p > .4$). To test whether the presence and congruence of co-occurring sounds had an influence on visual stimulus encoding we compared decoding accuracy in the three conditions containing visual stimuli (AV congruent, AV incongruent, V) for V1-3. Decoding performance did not differ significantly between conditions in V1 ($F_{(2,28)} = 0.46$, $p = .64$, *n.s.*). However, the presence and congruence of sounds had a significant effect on decoding performance in area V2 ($F_{(2,28)} = 7.17$, $p = .003$) and there was a non-significant trend for such an effect in area V3 ($F_{(2,28)} = 2.12$, $p = .14$, *n.s.*). Post-hoc t-tests revealed

that stimulus decoding from activity patterns in area V2 was significantly worse in the AV incongruent condition compared to both, decoding in the AV congruent ($t_{(14)}$ = 3.29, p = .005) and V ($t_{(14)}$ = 3.46, p = .004) conditions. Pattern decoding from area V3 was significantly worse for the AV incongruent condition compared to the V condition ($t_{(14)}$ = 2.15, p = .049).

To further investigate the effect of sounds on stimulus decoding from activation patterns in V1-3 I compared the reliability and similarity of stimulus-evoked patterns (cf. Methods for details). There was no detectable influence of sounds on pattern similarity in V1-3 (V1: F(2,28) = 0.762, p = .476, *n.s.*, V2: F(2,28) = 1.069, p = .357, *n.s.*, V3: F(2,28) = 1.815, p = .181, *n.s.*; cf. **Figure 5-2 d)**). However, pattern reliability was significantly affected by the presence of sounds in V2 and V3 (V1: F(2,28) = 2.013, p = .152, *n.s.*, V2: F(1.4,28, Greenhouse-Geisser corrected) = 6.647, p = .011, V3: F(2,28)=5.133 p=.013; cf. **Figure 5-2 c)**).) Post-hoc paired t-tests revealed that pattern reliability in V2 was significantly reduced in the AV incongruent condition, compared to both the AV congruent condition ($t_{(14)}$ = -2.376, p=.032) and the V condition ($t_{(14)}$ = -5.406, p < .0001). Pattern reliability in V3 was significantly reduced in the AV incongruent condition, compared to the V condition ($t_{(14)}$ = -3.004, p = .010).

The current study was limited to investigating multisensory modulation of pattern discriminability in early visual cortices. It would have been interesting to compare this to similar modulations in early auditory cortex. However, auditory pattern decoding from BOLD signals typically has much lower accuracies than visual pattern decoding and appears to require

high spatial resolution MRI sequences (e.g. Formisano et al., 2008; Staeren et al., 2009). Nevertheless, for completeness I also extracted patterns of BOLD signals from bilateral anterior transversal temporal gyri (Destrieux, Fischl, Dale, & Halgren, 2010) and tried to classify them. Stimulus decoding was generally unsuccessful for this data and did not improve even when using a more lenient anatomical criterion (including the whole of the superior temporal gyrus and plane). I conclude that an investigation of primary auditory cortex similar to my visual cortex analysis would rely on high-resolution scans and adequate functional localizers, ideally tonotopic-mapping.

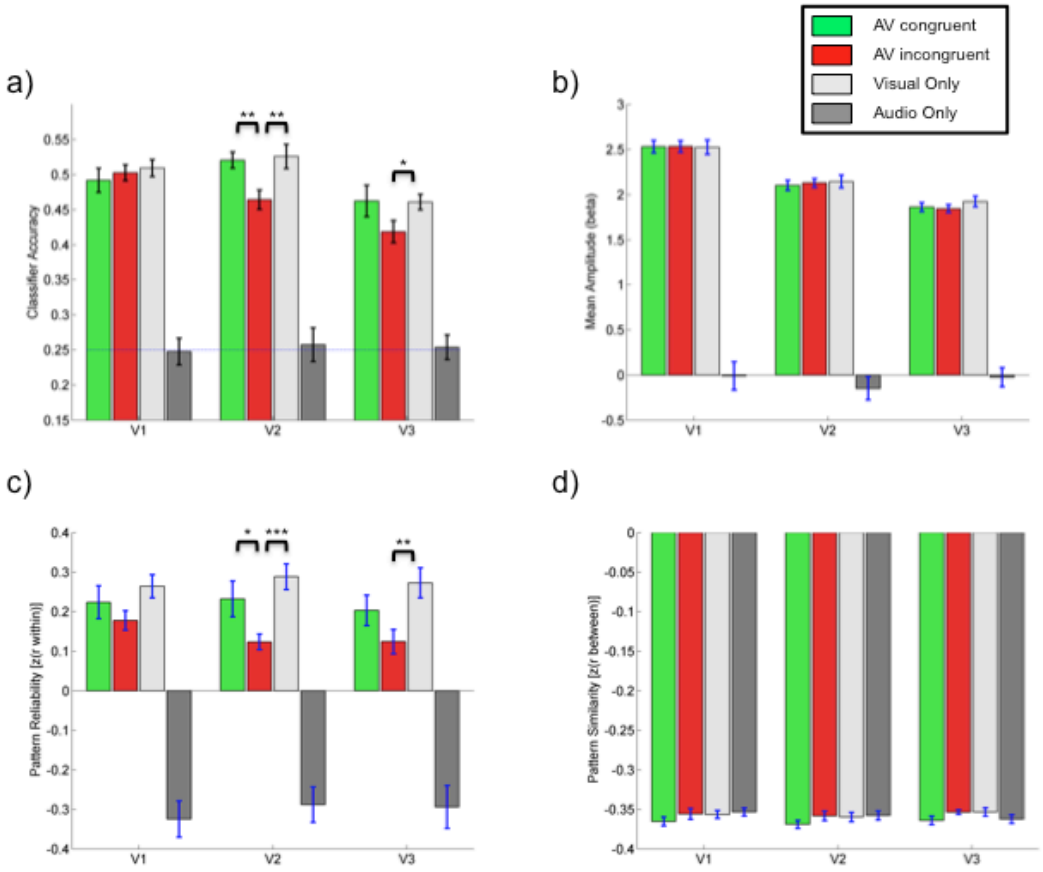

Figure 5-2 Results for Regions of Interest (ROIs). Results for areas V1-3 are shown as bar plots. Bar colors indicate conditions: audiovisual congruent in green, audiovisual incongruent in red, visual only in light grey and audio only in dark grey. Error bars indicate the standard error of the mean adjusted for repeated measurements (Morey, 2008). **a)** Classification accuracies for 4-way classification using linear support vector machines (see Methods for details). The dashed line indicates chance level (.25). Stars indicate significantly different decoding accuracies between conditions involving visual stimulation (as indicated by paired *t*-tests, see Results for details of respective ANOVAs; * $p < .05$, ** $p < .01$). **b)** Mean signal amplitudes estimated by the GLM. Note that amplitudes were not significantly different between conditions involving visual stimulation in any of the regions of interest. Note that beta maps used for this analysis were not mean corrected (see Methods for details). **c)** Pattern reliability as indicated by means of Fischer z-transformed correlation coefficients between patterns for a given stimulus in odd and even runs (see Methods for details). Stars indicate significantly different pattern reliabilities between conditions involving visual stimulation (as indicated by paired *t*-tests, see Results section for details of respective ANOVAs; * $p < .05$, ** $p < .01$, *** $p < .001$). **d)** Pattern similarity as indicated by means of Fischer z-transformed correlation coefficients between patterns for different stimuli (see Methods for details). Note that pattern similarities were not significantly different between conditions involving visual stimulation in any of the regions of interest. Patterns are negatively correlated because they were mean corrected across stimuli within each condition (see Methods for details).

5.3.2.2 Searchlight results

I tested three contrasts: AV congruent – AV incongruent, AV congruent – V and V – AV incongruent (see Methods for details.).

The AV congruent – AV incongruent contrast yielded no significant clusters at the corrected threshold. The AV congruent – V contrast revealed two significant clusters in the bilateral superior temporal gyri (FWE corrected $p < .05$). Both clusters included early auditory cortex and part of the superior temporal gyrus (including TE 1.0, 1.2 and 3) and the right cluster extended in anterior direction to the temporal pole (cf. **Table 5-1** and **Figure 5-3 a)**). The V – AV incongruent contrast yielded two significant clusters in visual cortex (FWE corrected $p < .05$). The first cluster spanned part of the bilateral calcarine gyrus near the occipital pole, including parts of Brodmann area 17 and 18. The second cluster was located in the left lateral inferior occipital gyrus and coincided with the location reported for areas LO1/2 (Larsson & Heeger, 2006). See **Table 5-1** and **Figure 5-3 b)**).

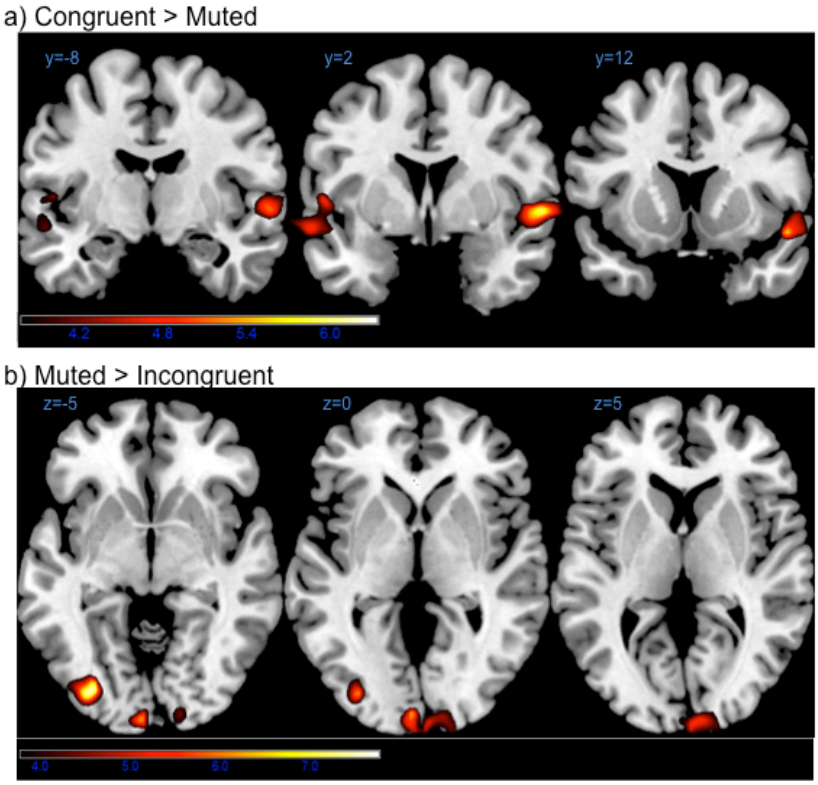

Figure 5-3 Results for Whole Brain Searchlight Analysis. Heat maps for searchlight contrasts overlaid on coronal and axial slices of a T1-weighted anatomical image ('Colin 27', (Holmes et al., 1998)). Searchlight maps indicating local pattern discriminability for each condition were normalized and contrasted on the second level (see Methods for details). Color coding for *t*-values is indicated by color bars at the bottom of **a)** and **b)**. Please note that the contrast between the audiovisual incongruent and congruent conditions was tested as well but yielded no significant results. Note that contrasts are directed and that contrasts of opposite direction yielded no significant results. **a)** Increased pattern discriminability for the audio-visual congruent condition as compared with the visual only condition in bilateral superior temporal gyrus (see Table 5.1 and Results for details). **b)** Increased pattern discriminability for the visual only condition as compared with the audio-visual incongruent condition in left lateral occipital area and the

Table 5-1 Significant searchlight clusters. Details of clusters where decoding accuracy was significantly different between conditions. Coordinates of peak voxels are in MNI space, cluster size is in voxels and *p*-values are whole brain FWE corrected at cluster level, *t*-values correspond to peak voxels. Anatomical labels refer to the Juelich Histological atlas. See Methods for details.

Contrast	*p* value	cluster size	*t*-value	peak voxel	label
AV congruent - V	<.001	861	6.33	[62 -2 0] [50 16 -12] [62 20 -12]	r Superior Temporal Gyrus r Temporal Pole (not assigned)
	.006	408	5.40	[-56 -2 8] [-52 4 2] [-60 6 -10]	l Superior Temporal Gyrus l Rolandic operculum (mot assigned)
V – AV incongruent	<.001	699	6.93	[-32 -82 -4]	l inferior Occipital Gyrus
	<.022	303	7.75	[-4 -94 -2]	l Calcarine Bank

5.3.3 Univariate fMRI results

To test where in the brain auditory context modulated the amplitude of the signal evoked by my stimuli (as opposed to information carried), I employed a univariate whole brain analysis. I tested the same three contrasts tested in the searchlight analysis: AV congruent – AV incongruent, AV congruent – V and V – AV incongruent (see Materials and Methods for details).

The AV congruent – AV incongruent contrast yielded no significant results. The AV congruent – V contrast yielded two significant clusters in the bilateral superior temporal gyri (FWE corrected $p < .05$). Both clusters

included early auditory cortex (including TE 1.0, 1.1, 1.2 and 3) and the right cluster extended in anterior direction to the temporal pole (cf. **Table 5-2** and **Figure 5-4 a**), note the similarity to the corresponding searchlight contrast). The V-AV incongruent contrast yielded two similar clusters of significantly greater activation for the AV incongruent condition (i.e. the one including auditory stimulation). These clusters again spanned almost the whole of bilateral superior temporal gyri, including early auditory cortex (cf. **Table 5-2** and **Figure 5-4 b**).

For a more direct comparison between univariate contrasts and the multivariate analysis we also tested for univariate effects in the retinotopically defined ROIs of each participant. For this contrast we averaged the voxel responses (betas) for each participant and condition across the whole of the respective ROI (cf. **Figure 5-2 b**). Response amplitudes did not differ significantly between the three conditions involving visual stimuli in all three ROIs (V1: $F_{(2,28)} = 0.01$, $p = .99$, *n.s.*; V2: $F_{(2,28)} = 0.25$, $p = .78$, *n.s.*; V3: $F_{(2,28)} = 1.12$, $p = .34$).

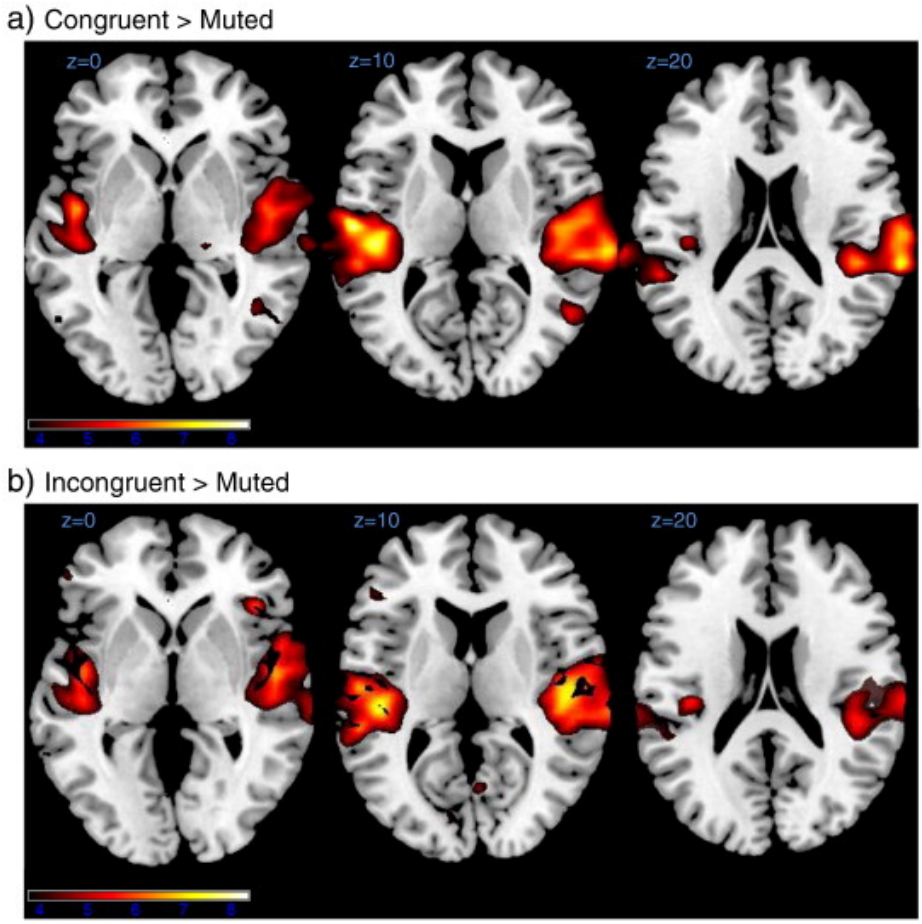

Figure 5-4 Results for Whole Brain Univariate Analysis. Heat maps indicating differences in signal amplitude between conditions overlaid on axial slices of a T1-weighted anatomical image ('Colin 27', (Holmes et al., 1998)). Color coding for t-values is indicated by the color bar at the bottom. See results and table 2 for details. Note that contrasts are directed and that contrasts of opposite direction yielded no significant results. **a)** Increased signal amplitude for the audio-visual congruent condition as compared with the visual only condition in bilateral superior temporal gyri. **b)** Increased signal amplitude for the audio-visual incongruent condition as compared with the visual only condition in bilateral superior temporal gyri.

Table 5-2 Significant clusters for the univariate Analysis. Details of clusters where BOLD signal intensities were significantly different between conditions. Coordinates of peak voxels are in MNI space, cluster size is in voxels and *p*-values are whole brain FWE corrected at cluster level, *t*-values correspond to peak voxels. Anatomical labels refer to the Juelich Histological atlas. See Methods for details.

Contrast	*p* value	cluster size	*t*-value	peak voxels	labels
AV congruent -V	<.001	1392	8.32	[57 -31 13]	r Superior Temporal Gyrus
			7.81	[69 -22 16]	"
			7.69	[54 -7 -8]	"
	<.001	900	8.26	[-57 -16 10]	l Superior Temporal Gyrus
			7.76	[-48 -25 10]	"
			7.05	[-42 -19 13]	l Rolandic Operculum

5.4 Discussion

5.4.1 Summary of results

I presented participants with naturalistic, dynamic audiovisual stimuli while they performed an incidental fixation task. Replicating previous studies (e.g. (Nishimoto et al., 2011)), I could decode stimulus identity from spatial patterns of BOLD signals in retinotopic cortices well above chance. More specifically, I could decode stimulus identity significantly better than chance from BOLD patterns in V1-3 (separately) for all conditions containing visual stimuli (AV congruent, AV incongruent and V), but not for the audio only (A) condition.

There were no detectable differences in mean amplitudes of BOLD signals evoked in V1-3 for the AV congruent, AV incongruent and V conditions. However, and most importantly, decoding accuracy varied significantly with the presence and congruence of sounds in V2 and

somewhat in V3. Decoding accuracy for patterns in V2 was worse for the AV incongruent condition compared to both, the V and AV congruent condition. Decoding accuracy in V3 was worse for the AV incongruent compared to the V condition. Worsening of local decoding accuracies for the AV incongruent (compared to V) condition was confirmed and extended to area LO (and possibly V1) by searchlight analyses.

Significantly worse decoding for the AV incongruent condition in V2 (compared to the AV congruent and V conditions) was associated with reduced inter-trial reliability of patterns for a given stimulus in this condition (again, in comparison to the AV congruent and V conditions). In V3 reduced decoding accuracy for the AV incongruent condition relative to the V condition went along with reduced inter-trial reliability for the same comparison. In contrast to the reliability of *intra*-stimulus patterns, no significant modulation of *inter*-stimulus pattern similarity could be found.

5.4.2 Modulation of pattern discriminability

The results of the experiment presented in this chapter demonstrate modulation of stimulus evoked pattern discriminability as a consequence of multisensory interactions in human early retinotopic cortex. They are in accord with and extend recent findings in macaque primary auditory cortex (Kayser et al., 2010) and superior temporal sulcus (Dahl et al., 2010). Notably, I observed these modulations in early visual cortex using high-contrast visual stimuli that covered only central parts of the visual field (<10 degrees eccentricity). The data suggest that this effect reflected modulations of inter-trial reliability of neural activation patterns for a given stimulus, i.e.

the average multivariate mean for a given stimulus was not shifted, but the trial-by-trial scatter around this mean depended on multisensory context. This is also in line with the findings of Kayser et al. (2010) and Dahl et al. (2010) who found multisensory modulation of inter-trial reliability for temporal spike patterns of single neurons.

Note that I could not discriminate BOLD signal patterns in visual cortex evoked by purely auditory stimuli. This contrasts with the findings that auditory motion direction can be decoded from lateral occipital cortex (Alink, Euler, Kriegeskorte, Singer, & Kohler, 2012) and visual stimulus identity can be decoded from early auditory cortex (Hsieh, Colas, & Kanwisher, 2012; Meyer et al., 2010). A possible explanation for this difference is that such effects rely on top-down attention or even cross-modally evoked imagery (Hsieh et al., 2012; Meyer et al., 2010). It is possible that this kind of effect was prohibited or attenuated by the fixation task I used. Alternatively, it is possible that only certain types of auditory signal such as those associated with motion can be decoded from visual cortex.

Interestingly modulations of BOLD pattern discriminability in visual cortices were not accompanied by overall amplitude modulations in the experiment presented here. This differs from the results of previous fMRI studies that found increased univariate signals in early sensory areas for audiovisual concurrent compared to purely visual stimulation (Martuzzi et al., 2007; Noesselt et al., 2007; Watkins et al., 2006). This difference might reflect the fact that these earlier studies used transient, periliminal or low contrast stimuli while I used naturalistic stimuli in the current experiment. Also, Kayser et al. (2010) and Dahl et al. (2010) found some net *reduction* in

firing rate for bimodal vs. unimodal stimulation. However, the present V condition differed from their design: in the current experiment it was not truly unimodal because scanner noise was present throughout the experiment. Whatever the reasons for the dissociation between modulation of overall amplitude and pattern discriminability in the present work, it renders my results important in the context of the debate about criteria for multisensory interactions. These usually concern different types of amplitude modulation and the question which of them qualify as 'multisensory' (e.g. (Beauchamp, 2005)). The current results demonstrate multisensory interactions in the absence of *any* detectable modulation of net amplitude. Furthermore, one might argue that, in the context of naturalistic stimuli, modulation of pattern discriminability may be the most relevant effect of multisensory interactions. Recently, it has been argued that the role of primary sensory cortices in audio-visual integration might be limited to low level stimulus features and transient stimuli (Giani et al., 2012; Werner & Noppeney, 2010). The basis for this argument is the observed insensitivity of the (univariate) BOLD signal amplitude in primary auditory cortex to higher order stimulus congruence (Werner & Noppeney, 2010) and the absence of cross-modulation frequencies for audio-visual steady-state responses in MEG ((Giani et al., 2012); note that the latter method does not allow the presentation of audio-visual congruent stimuli). The current results suggest the null results in these studies could reflect an insensitivity of the analysis methods used to detect modulations of the encoded stimulus information (like pattern discriminability or pattern reliability). This

underscores the need for further research to clarify the exact role of primary sensory cortices in audiovisual stimulus integration.

5.4.3 Potential Mechanisms modulating audiovisual pattern discriminability

How do sounds affect the reliability of early visual cortex signals? Most likely this effect rests on subthreshold modulation of visual neurons, rather than on classical bimodal neurons. Bimodal neurons in early visual cortex seem to be restricted to the far periphery of visual space (which I did not stimulate here) whereas subthreshold modulation also affects more central representations (Allman & Meredith, 2007). Furthermore, multisensory modulation of spike train discriminability is found for subthreshold modulation of visual neurons (Dahl et al., 2010). One could speculate that such subthreshold modulation in turn could be mediated via phase alignment of ongoing oscillations (Lakatos et al., 2007; Naue et al., 2011; Romei et al., 2012). Some results from a recent MEG study are of particular interest (Luo et al., 2010), showing that accuracy of decoding video stimuli from phase patterns of occipital channels depends on audiovisual congruency. Furthermore, in that MEG study the trial-by-trial phase coherence (i.e. reliability) for a given video stimulus was affected by audiovisual congruency as well. It has been proposed that temporal profiles of neural activity in different primary sensory areas can work as oscillatory attractors on each other, effectively yielding an ongoing modulation of excitability (Lakatos et al., 2009; Schroeder et al., 2008). This could serve to minimize temporal uncertainty (Friston, 2009) and would be very similar to

what was proposed as an early theory of 'dynamic attention' (Jones, 1976; Large & Jones, 1999). Note, that for my design such effects would likely be stimulus driven, rather than top-down controlled – participants were engaged in a fixation task and had no incentive to concentrate on the dynamic stimuli in the background.

If temporal fine-tuning is indeed a mechanism behind my finding, it is interesting that MVPA was sensitive enough to pick it up despite the relatively coarse temporal resolution of fMRI and the fact that decoding rests on *spatial* patterns of activation. The studies by Kayser et al. (2010) and Dahl et al. (2010) investigated modulation of single unit firing rate variability. This could translate to BOLD pattern variability, if the variance of the net population amplitude in a voxel would be modulated in effect – or at least the variance of modulatory pre-synaptic activity contributing to the BOLD-signal (Cardoso et al., 2012; Friston, 2012).

5.4.4 Null results with regard to enhanced pattern discriminability and V1

My data did not show significant modulation of pattern discriminability in V1. For V2 and V3 they only showed reduced pattern discriminability in the AV incongruent condition, but no enhancement for the AV congruent condition. Null-results need to be interpreted cautiously for several reasons. In this case, there are additional, design-specific reasons to be cautious: Multisensory interactions are generally more likely for peripheral (e.g. (Allman & Meredith, 2007)) and degraded (e.g. (Ernst & Banks, 2002; Fetsch, Pouget, DeAngelis, & Angelaki, 2012)) stimuli.

However, the visual stimuli used here were naturalistic and had high contrast, while the sounds I used were unavoidably degraded due to simultaneous scanner noise associated with BOLD signal acquisition. Thus my design was suboptimal for evoking maximum cross-modal interaction effects and potentially biased towards detrimental effects on visual processing rather than enhancement. That said, one might expect audio-visual effects to be stronger in V2 than V1 if they rest on direct crosstalk with auditory cortex, because these connections seem to be much sparser in V1 than in V2 (Rockland & Ojima, 2003). Furthermore, Kayser et al. (2010) found enhancement of information representation in macaque A1 for AV congruent as well as for AV incongruent stimuli. However, Dahl et al. (2010) found only significant information degradation for visual neurons in the AV incongruent condition, but no significant enhancement for the AV congruent condition. In sum, it might be possible that the signal to noise ratio (SNR) of early visual responses is close to ceiling for naturalistic stimuli, and thus early auditory responses are more likely to gain from multisensory interactions. Future studies could parametrically vary the SNR of visual stimuli (or possibly both modalities) to shed further light on this question.

5.4.5 Possible Sources of multisensory interactions

My data provide information about the effects of multisensory interactions in V1-3, but not about their source(s). The multisensory effects I observed could be mediated by feedback connections from multisensory cortices, by feed-forward connections from the superior colliculus and/or by direct connections between primary sensory areas (cf. (Driver &

Noesselt, 2008) and (Klemen & Chambers, 2012) for an overview). In humans, analyses of functional connectivity could provide hints regarding these possibilities (e.g. psycho-physiological interactions (PPI) (Friston et al., 1997)). Unfortunately, however, the optimal design requirements for MVPA are very different from those for connectivity analyses (e.g. fast event related designs to acquire many pattern examples for MVPA vs. longer task blocks for PPI). Future studies could try to combine both analysis techniques by applying both kinds of designs in one sample. This would allow testing for correlations between the individual strength of modulation with regard to information representation and with regard to connectivity.

5.4.6 Conclusion

Multisensory interactions affect human visual cortex processing from its earliest stages. For naturalistic stimuli, these interactions can be restricted to reliability modulations of fine-grained patterns and thus go undetected by common univariate analyses. This calls into question the exclusivity of criteria for multisensory interactions involving net amplitude modulation. The purpose of pattern discriminability modulations is likely to enhance encoding reliability (esp. for weak stimuli), but further research is needed.

This was the final chapter on auditory modulation of visual perception and visual system activity. In this chapter I probed how general aspects of information representation in V1-3 are modulated by auditory input (BOLD pattern reliability and discriminability). In the next chapter I will ask whether a specific aspect of visual representations in V1-3 is

modulated by perceptual load, namely the spatial resolution of stimulus representations.

6 Does perceptual load modulate the spatial tuning of population responses in V1-3?

6.1 Introduction

Load theory (Lavie, 1995, 2005) proposes that increasing processing load associated with an attended target will suppress perceptual processing of distractors due to progressive exhaustion of fixed processing capacity. Consistent with this, high perceptual load reduces distractor related interference (e.g. Lavie, 1995), visual sensitivity (e.g. (Carmel et al., 2011)), adaptation (e.g. (Rees et al., 1997)), visual cortex excitability (Muggleton et al., 2008) and overall levels of activity in human visual cortex ((Schwartz et al., 2005); also see Chapter 1.3.2.2). However it is unclear whether perceptual load also affects other fundamental properties of neuronal populations in human visual cortex, such as their spatial tuning.

Attention has been shown to affect the spatial tuning of single neurons in paradigms manipulating covert spatial attention (e.g. (Womelsdorf et al., 2006, 2008)). These studies show that neurons in the extrastriate cortex of monkeys can shift their receptive fields towards the attended location and shrink around attended stimuli within the receptive field (see Chapter 1.3.3.2 for a more detailed review). Studies measuring the effects of covert spatial attention on visual acuity (e.g. (Montagna et al., 2009)), texture segmentation performance (e.g. (Yeshurun & Carrasco, 1998)) and visual appearance (e.g. (Abrams et al.,

2010)) show that spatial attention also affects spatial aspects of stimulus representation in humans. Spatial attention seems to enhance the spatial resolution of the representation for attended locations (see Chapter 1.3.3 for more details).

Finally, slight differences in stimulus location are better reflected in the evoked pattern of BOLD responses in human V1-4 when the stimuli are attended (Fischer & Whitney, 2009). Thus spatial attention seems to refine the spatial tuning of the population response.

Here, I sought to test whether perceptual load at fixation can modulate the spatial tuning of neural population responses representing the surrounding visual field. I used fMRI and population receptive field (pRF) mapping (Dumoulin & Wandell, 2008) to estimate spatial tuning functions of neuronal populations in human V1-3. A population receptive field refers to the location and spatial extent of positions in the visual field where stimulation evokes responses at a corresponding location (voxel) in visual cortex; they provide a non-invasive measure of the spatial tuning of neuronal populations within human visual cortices (see Chapter 2.5.5.3 for more details). At the same time I manipulated perceptual load at central fixation. Participants performed a task of either high or low perceptual load on identical streams of stimuli while task-irrelevant mapping stimuli traversed the visual field (**Figure 6-1**; details in Methods section). This enabled me to compare the location and size of pRFs under high vs. low load.

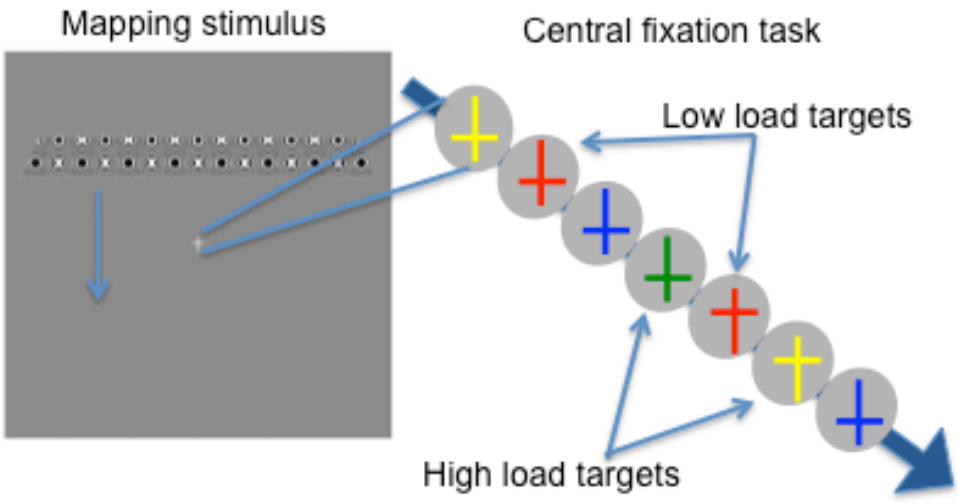

Figure 6-1 Stimuli and task. The left hand side of the figure shows an example screen from the stimulus sequence. The blue arrow and lines only serve illustration purposes and were not shown. While participants solved a central fixation task, task-irrelevant bar-type mapping stimuli traversed the surrounding visual field (traveling downwards towards fixation in the example, as indicated by the blue arrow). The right hand side of the figure is a magnification of the central fixation stimulus and shows a time series of stimuli in the fixation task, indicating targets in either condition. The fixation task required detection of targets in a rapid series of colored upright or inverted crosses (Schwartz et al., 2005). Targets were defined either based on color alone (low load; red crosses) or by a conjunction of color and orientation (high load; upright yellow and inverted green cross). Target frequency and stimulus streams were identical between conditions (see Methods section below for details).

6.2 Methods

6.2.1 Participants

Twenty-seven healthy participants with normal or corrected-to-normal visual acuity completed the experiment. Participants were recruited from the University College London (UCL) participant pool and gave written informed consent to take part in the study, which was approved by the UCL ethics committee. Data from one participant were excluded because reconstruction of the cortical surface failed, leaving 26 participants in the

main analysis (mean age, 25 yrs, SD, 5 yrs; 14 females; 2 left handed). One participant took part in both versions of the experiment (conditions alternating between vs. within runs, see below), thus a total of 27 datasets were entered into the main analysis. Note that excluding the 2nd dataset of the participant scanned twice did not change the results.

6.2.2 Stimuli

Stimulus streams presented at central fixation were similar to those used in Schwartz *et al*. (Schwartz et al., 2005) and consisted of a rapid serial visual presentation (RSVP) of colored, upright or inverted crosses (height: 0.7° visual angle, with the horizontal bar centered at a vertical distance of 0.28° visual angle from the upper or lower tip, respectively; colors: red, green, yellow, blue, black, white, light blue, purple, cyan, pink, orange, violet, brown) (c.f. **Figure 6-1**). Each cross was presented on a grey background at the center of the display for 500 ms with a gap of 250 ms between successive crosses. Targets in the low load condition were defined as red crosses (regardless of orientation); in the high load condition targets were defined as upright yellow and inverted green crosses. Stimuli did not differ between conditions and the proportion of targets among all stimuli was ~7.5% for either condition.

Task-irrelevant mapping stimuli consisted of bars containing a dynamic, high contrast black (1.3 cd/m^2) and white (1997.5 cd/m^2) non-Cartesian grating (775.1 cd/m^2). Bars were ~1.5° visual angle wide and traversed a circular area subtending 9° visual angle from fixation. They moved in horizontal or vertical sweeps of 24 steps (step duration 2.55 s,

corresponding to one TR) and spared a circular fixation aperture at the center of the display, which was ~1.4° visual angle wide. The fixation aperture contained either the load stimuli or a fixation dot, which was ~0.2° visual angle wide. The background of the display was a uniform grey (547.5 cd/m^2). At all times, a subtly darker grey static 'spider web' was superimposed onto the entire display (mapping stimuli and background) to aid fixation compliance.

For each participant I acquired additional runs to estimate the individual, condition specific hemodynamic response function (HRF). I used sparse photic bursts for this, which filled the whole of the mapped area (stimulus duration: 2.55 s; inter-stimulus interval: 28.05 s). The stimulus contained the same pattern as the mapping bars.

Participants viewed stimuli via a mirror mounted at the head coil at a viewing distance of ~61 cm and at a resolution of 1024x768 pixels. All stimuli were programmed and presented in MATLAB (Mathworks, Ltd.) using the Psychophysics Toolbox 3 extension(Brainard, 1997; Pelli, 1997) (http://psychtoolbox.org).

6.2.3 Procedure

Each participant was familiarized with the load task outside the scanner and completed 4-8 mapping runs and two runs for HRF estimation. Load conditions alternated between runs for 14 participants (mapping runs of 148 volumes (~6 minutes), HRF runs of 104 volumes (~4.5 minutes) and within runs for 13 participants (mapping runs of 196 volumes (~8 minutes), HRF runs of 172 volumes (~7 minutes)).

Participants for which conditions alternated *between* runs (sample A) were notified of the upcoming condition by an instruction screen at the beginning of each run and the order of runs was counterbalanced between participants. Each mapping run was divided into 6 epochs of 24 TRs (or 61.2 s) and participants solved the ongoing load task throughout. The mapping stimulus traversed the display during 4 of the epochs, with each epoch corresponding to one cardinal sweep direction (order of directions pseudo-randomized but constant across load conditions). The third and sixth epochs were blank epochs containing no mapping stimuli. After four mapping runs, participants completed two HRF runs. During HRF runs, participants solved the ongoing load task while 10 photic bursts per run were presented in the periphery (see above, pRF mapping stimuli).

Participants for which conditions alternated *within* runs (sample B) were notified of the condition by a brief cue presented at the center of the screen (both targets next to each other). Mapping runs were divided into 8 epochs of 24 TRs, with load conditions alternating after 4 epochs. Two mapping epochs (corresponding to two cardinal sweep directions) were always followed by two blank epochs. The load task was ongoing during the first of these blanks, while the second blank was a rest period during which participants were required to fixate a dot at the center of the display (serving as a common baseline for both load conditions). After three mapping runs, participants completed two HRF runs with 5 photic bursts per load condition. After the first 5 photic bursts participants carried on solving the load task for 12 blank TRs (or 30.6 s), followed by a rest period

of 24 TRs (60.12 s). Then the task resumed (with the other load condition) for another 5 photic bursts and 12 blank TRs.

To assess fixation compliance I used an EyeLink 1000 MRI compatible eyetracker (http://www.sr-research.com/), tracking gaze position of the left eye. Due to technical problems and a restricted field of view for the eyetracker I could only collect eye data for 16 participants.

6.2.4 Image acquisition and pre-processing

All functional and structural scans were obtained with a Tim Trio 3T scanner (Siemens Medical Systems, Erlangen, Germany), using a 32-channel head coil. However, the front part of the head coil was removed for functional scans, leaving 20 effective channels (this way restrictions of participants' field of view were minimized). Functional images for the main experiment were acquired with a gradient echo planar imaging (EPI) sequence (2.3 mm isotropic resolution, matrix size 96 x 96, 30 transverse slices per volume, acquired in interleaved order and centered on the occipital cortex; slice acquisition time 85 ms, TE 37 ms, TR 2.55 s). I obtained 148 volumes per mapping run and 124 volumes per HRF run (196 and 172 volumes, respectively, for mapping and HRF runs including both load conditions; including four dummy volumes at the beginning of each run). In between mapping and HRF runs I acquired B0 field maps to correct for geometric distortions in the functional images caused by heterogeneities in the B0 magnetic field (double-echo FLASH sequence with a short TE of 10 ms and a long TE of 12.46 ms, 3x3x2 mm, 1 mm gap). Finally, I acquired two T1-weighted structural images of each participant. The first structural image

was obtained with the front part of the head coil removed, using an MPRAGE sequence (1 mm isotropic resolution, 176 sagittal slices, matrix size 256 x 215, TE 2.97 ms, TR 1900 ms). For the second structural image I used the full 32-channel head coil with a 3D MDEFT sequence ((Deichmann et al., 2004); 1 mm isotropic resolution, 176 sagittal partitions, matrix size 256 x 240, TE 2.48 ms, TR 7.92 ms, TI 910 ms).

All image files were converted to NIfTI format and pre-processed using SPM 8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). The first four volumes for each run were discarded to allow for the T1 signal to reach steady state. The remaining functional images were mean bias corrected, realigned, unwarped (using voxel displacement maps generated from the fieldmaps(Hutton et al., 2002)), co-registered (with the respective anatomical MDEFT scan for each participant, using the MPRAGE scan as an intermediate step) and smoothed with a 4 mm Gaussian kernel. The anatomical MDEFT scan was used to reconstruct the cortical surface with FreeSurfer (http://surfer.nmr.mgh.harvard.edu) and the functional time series were projected onto the surface, detrended and z-normalized for each run and vertex. Finally, runs containing both load conditions were split, so that all data could be separated according to condition.

6.2.5 Data analysis

Data analysis was conducted using FreeSurfer (http://surfer.nmr.mgh.harvard.edu) and MATLAB (Mathworks, Ltd.), including SPM 8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) and a custom MATLAB toolbox for population receptive field analysis and

transforming data between volume and surface space. All data analyses were restricted to a mask of the posterior part of the brain, including occipital and inferior parietal cortex.

To estimate the individual, hemisphere and condition specific hemodynamic response function (HRF), I first identified visually responsive vertices within the mask (defined as vertices with an average positive response > 1 standard error for the first five volumes after each photic burst in an HRF run) and averaged the signal measured for all 10 photic bursts per scan run. I then fitted a double gamma function (Friston, Fletcher, et al., 1998) with four free parameters to the average stimulus evoked response of all visually responsive vertices. The fitted parameters corresponded to the latency of the HRF response and undershoot as well as their amplitudes (one parameter for the ratio of peak and undershoot amplitudes plus an absolute scale factor). To compare the amplitude of stimulus evoked responses in the periphery between load conditions I compared condition and hemisphere specific HRF amplitudes for those runs containing both conditions and a common baseline (n=26 hemispheres). I used paired t-tests to compare fitted parameters as well as raw amplitudes between conditions (two-tailed .05 significance level; for raw data the peak was defined as the maximum response obtained in either condition).

Population receptive field (pRF) modelling was based on the assumption of symmetric two-dimensional Gaussian pRFs and data from the two load conditions were fit independently to compare the two resulting sets of model parameters. Model fitting was performed in a similar fashion as described by (Dumoulin & Wandell, 2008). The three pRF parameters (x

and y coordinates of center position and the standard deviation, σ) were fitted in a two stage procedure for each vertex. Model predictions were always based on the stimulus time course (coding spatial positions as stimulated, or not, for a given point in time) and spatial sensitivity according to the assumed pRF parameters. To compare model predictions with empirical BOLD time courses, predictions were convolved with the hemisphere and condition specific HRF estimates (see above). A first coarse fit consisted of an exhaustive grid search for the set of parameters providing the highest correlation between empirical and predicted time courses (grid size 15x15x34 for center positions and σ, respectively). The obtained parameters then formed the initial values for a subsequent fine fit, aiming to minimize model prediction error using a simplex method (Lagarias, Reeds, Wright, & Wright, 1998; Nelder & Mead, 1965). This step also included a scaling factor, β, to estimate the signal strength. The resulting parameter maps were smoothed with a surface based kernel of 5 mm FWHM.

I delineated retinotopic regions of interest (V1-3, V3A/B, and IPS0/1) based on data from the low load condition (but area boundaries were checked and found to be consistent between conditions). Centre position coordinates of vertices were transformed into polar angle and eccentricity, color coded and projected as maps on the inflated cortical surface. Boundaries of V1-3 were drawn according to standard procedures (Sereno et al., 1995) at meridian mirror reversals of polar angle. Dorsal areas IPS01 could only be delineated for a subset of 24 hemispheres and data within dorsal areas was patchy for most hemispheres. Definitions of dorsal retinotopic areas followed Wandell et al. (Wandell et al., 2007): IPS0/1 was

defined as the area extending along the intraparietal sulcus from the dorsal boundary of V3A/B to the next representation of the upper vertical meridian (with V3A/B being defined as the area anterior and superior of V3, bordering on a representation of the upper vertical meridian dorsally).

I compared pRF eccentricity and size (σ) between conditions. Eccentricity was defined as the distance between fixation and the estimated pRF center position. Vertices with a model fit of R^2<.05 in either of the conditions were not taken into account for statistical comparisons. I calculated differences on a vertex by vertex basis (i.e. parameters for a given vertex were compared between conditions) and as relative change compared to the size of the pRF under low load (absolute differences divided by σ). I binned results in eccentricity bands of half a degree visual angle, based on pRF center positions according to low load data. This analysis was restricted to eccentricities between one and seven degrees thus avoiding inner- and outermost pRFs, which were only partially mapped. The resulting differences (and absolute values) were averaged for each eccentricity band and hemisphere. I then applied an outlier criterion to data in each eccentricity band independently, removing data points further than 3.5 robustly estimated standard deviations from the mean (standard deviations estimated based on median absolute deviation (Hampel, 1974)). Note that the effects did not hinge on outlier removal and persisted when potential outliers were left in the analysis. Differences were tested against zero based on 2nd level one-sample t-tests (two-tailed .05 significance level).

Note that the forward modelling approach I used to estimate pRFs explicitly incorporated the hemodynamic response function (HRF).

Perceptual load could have an effect on the HRF that is unrelated to the spatial preference of the underlying neural populations (e.g. due to changes in neural amplitude (Schwartz et al., 2005), response latency and/or HRF shape). I aimed to control for potential HRF confounds by collecting additional data for each participant and estimating the visual cortex HRF on a hemisphere and condition specific basis (see above). For the main pRF analyses I used these condition and participant-specific HRFs. This also allowed me to explicitly test for effects of condition on BOLD amplitude by comparing HRF estimates between conditions (see **Figure 6-4**). To additionally test the effect of HRF estimates on pRF estimates I repeated the pRF analysis, this time swapping the estimated HRFs between conditions (see **Figure 6-6**).

Eye movement data were compared between conditions by calculating the average standard deviation (*S.D.*) of eye position across epochs in the mapping experiment (independently for horizontal and vertical axes and separately for each condition; **Figure 6-3**)). Note that simulations indicate pRF estimates can be biased by eye movements, but only if they are of considerable magnitude (c.f. Figure 6 in (Levin, Dumoulin, Winawer, Dougherty, & Wandell, 2010)). To further test whether participants were biased in their eye movements towards the stimulus, the same analysis was repeated for epochs with horizontal and vertical bar sweeps separately. This allowed me to calculate an index of bias in eye position towards the stimulus as

$$\text{Bias}=(S.D._{\text{vertical}}-S.D._{\text{horizontal}} \mid \text{vertical stimulus sweep})-(S.D._{\text{vertical}}-S.D._{\text{horizontal}} \mid \text{horizontal stimulus sweep})$$

6.3 Results

6.3.1 Behavioral data

6.3.1.1 Load task

Participants were significantly less sensitive (t_{25}=11.83, P<10^-11) and slower (t_{25}=15.75, P<10^-13) for detecting high vs. low load targets (**Figure 6-2**). This indicates a successful manipulation of perceptual load.

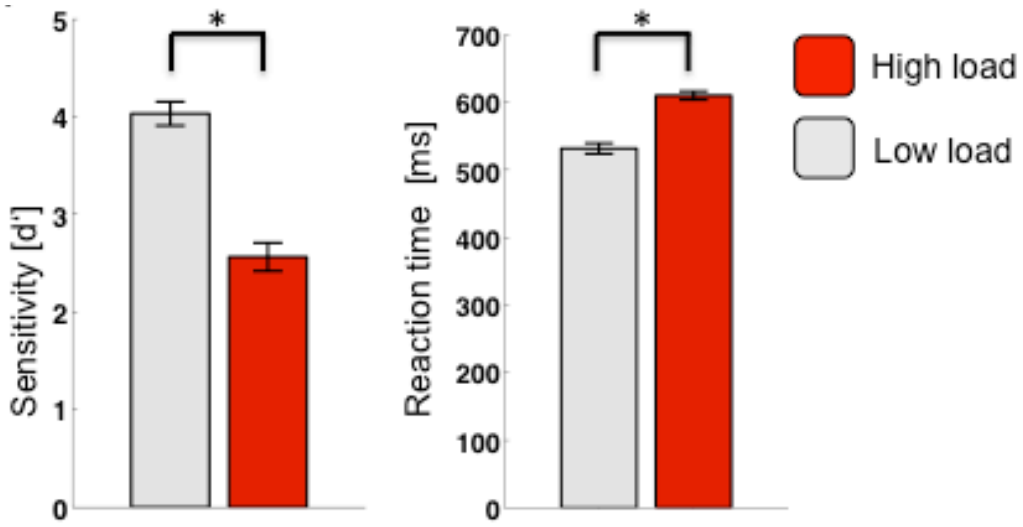

Figure 6-2 Behavioral data. The left panel shows participants sensitivity [d'] for target detection in either condition (color indicates condition as shown in the inset). Bars show mean sensitivity across participants, error bars indicate +/- one standard error of the mean (S.E.M). Participants were significantly less sensitive for targets in the high load condition. The right hand panel shows reaction times for hits by condition using the same color convention, bars indicate mean reaction time across participants, error bars indicate +/- one S.E.M. Participants were significantly slower when detecting high vs. low load targets.

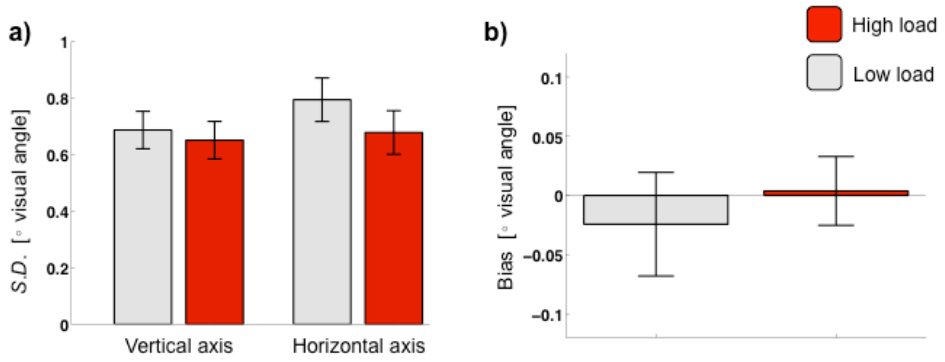

Figure 6-3 Eye movements. **a)** Variance in eye position by condition for the vertical and horizontal axes. Bars indicate the average standard deviation (S.D.) in eye position across participants and for the vertical (left hand side) and horizontal (right hand side) axes. S.D.s are expressed in degrees visual angle and were calculated for each stimulus sweep separately before averaging them for each participant. Error bars indicate +/- one standard error of the mean (S.E.M.). Participants showed a high degree of gaze stability (average SD < 1 degree visual angle for both conditions and axes). **b)** Eye movements towards mapping stimuli. To test whether participants were biased to move their eyes towards the mapping stimuli I calculated an index of eye movement bias towards the axis along which the stimulus travelled (see Methods section for details). Bars and error bars indicate the mean bias and +/- one S.E.M., respectively. Bias is expressed in degrees visual angle. Participants were not biased to follow the mapping stimulus with their gaze in either condition.

6.3.1.2 Gaze Behavior

Participants (n=16) showed a high degree of gaze stability with an average standard deviation of less than one degree visual angle per axes for both conditions (**Figure 6-3** a)). Participants' gaze was slightly more stable in the high load condition (vertical axis: $t(15)=3.14$, $P<0.01$; horizontal axis: $t(15)=1.69$, $P=0.11$, n.s.). Note that this is the opposite direction of what would explain the observed effects on pRF size estimates.

To test whether participants were biased to move their eyes towards the mapping stimuli I calculated an index of eye movement bias towards the axis along which the stimulus travelled (see Methods section for details). Participants were not biased to follow the mapping stimulus with their gaze

in either condition (**Figure 6-3** b); low load: t(15)=-0.56, P=0.59, n.s.; high load: t(15)=0.13, P=0.90, n.s.).

6.3.2 fMRI data

6.3.2.1 Hemodynamic response function

I explicitly compared HRFs between conditions for HRF runs containing both conditions and a common baseline (sample B; see Methods section for details). The peak amplitude of raw hemodynamic responses to the photic bursts was higher under low compared to high perceptual load for (t_{25}=3.04, P<.01; **Figure 6-4**). A similar, but non-significant trend was observed for the amplitudes of the fitted HRFs (t_{25}=1.91, P=.07). No significant difference was observed for any of the other fitted parameters (response latency: t_{25}=0.02, P=.99 undershoot latency: t_{25}=0.98, P=.34; peak/undershoot ratio: t_{25}=0.75, P=.46).

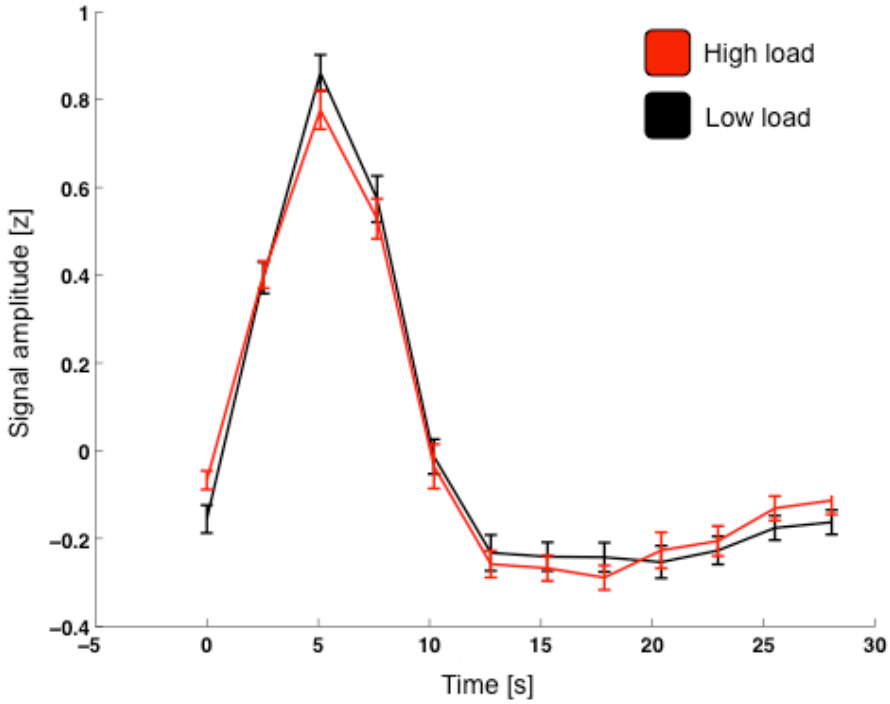

Figure 6-4 Comparison of hemodynamic responses to photic bursts across conditions for HRF runs containing both conditions (sample B). Data points (error bars) represent the mean +/- one SEM of z-normalized BOLD-signal amplitude across 26 hemispheres. There is one data point per TR after stimulus onset. Responses in the high load condition are shown in red; responses in the low load condition are shown in black. The peak amplitude of raw measurements to estimate the HRF was higher under low compared to high perceptual load for runs containing both conditions. A similar, but non-significant trend was observed for the amplitudes of the fitted HRFs (not shown; see Methods section for details).

6.3.2.2 Population receptive fields

6.3.2.2.1 Main results

High (versus low) perceptual load in the central task significantly affected the spatial tuning of early visual cortex for task-irrelevant stimuli presented in the surrounding visual field (**Figure 6-5**). I compared the size of pRFs with center positions from 1-7 degrees eccentricity by calculating the relative size difference between high vs. low load on a vertex by vertex

basis. pRF size significantly increased under high perceptual load across participants and hemispheres for V1 (mean increase=12.31% +/- s.e.m.=4.14%, t_{48}=2.97, P<0.01), V2 (mean increase =8.79% +/- s.e.m.=3.36%, t_{48}=2.61, P=0.01) and V3 (mean increase =10.95% +/- s.e.m.=3.61%, t_{49}=3.03, P<0.01). In both conditions and all three areas pRF size monotonically increased with eccentricity (**Figure 6-5** a)). However, from ~3-4 degrees eccentricity pRFs were significantly bigger for high perceptual load at fixation; early visual cortex representations of the surrounding region were blurred when the central fixation task was higher in perceptual load (**Figure 6-5** c)).

Apart from pRF size, perceptual load also affected pRF locations. I calculated shifts of pRF center positions (comparing high versus low load) on a vertex by vertex basis and expressed them relative to the respective pRF sizes. Average center position of pRFs became significantly more eccentric under high load in V1 (mean=5.92% +/- s.e.m.=2.45%, t_{45}=2.41, P<0.05) and V3 (mean=8.29% +/- s.e.m.=2.44%, t_{45}=3.40, P<0.01) with a similar trend in V2 (mean=3.81% +/- s.e.m.=2.71%, t_{51}=1.41, P=0.17; **Figure 6-5** d)). Like the blurring effect, this 'centrifugal' effect on pRFs was strongest from ~3-4 degrees eccentricity. However, it showed a steep decline after a peak at ~4-5 degrees).

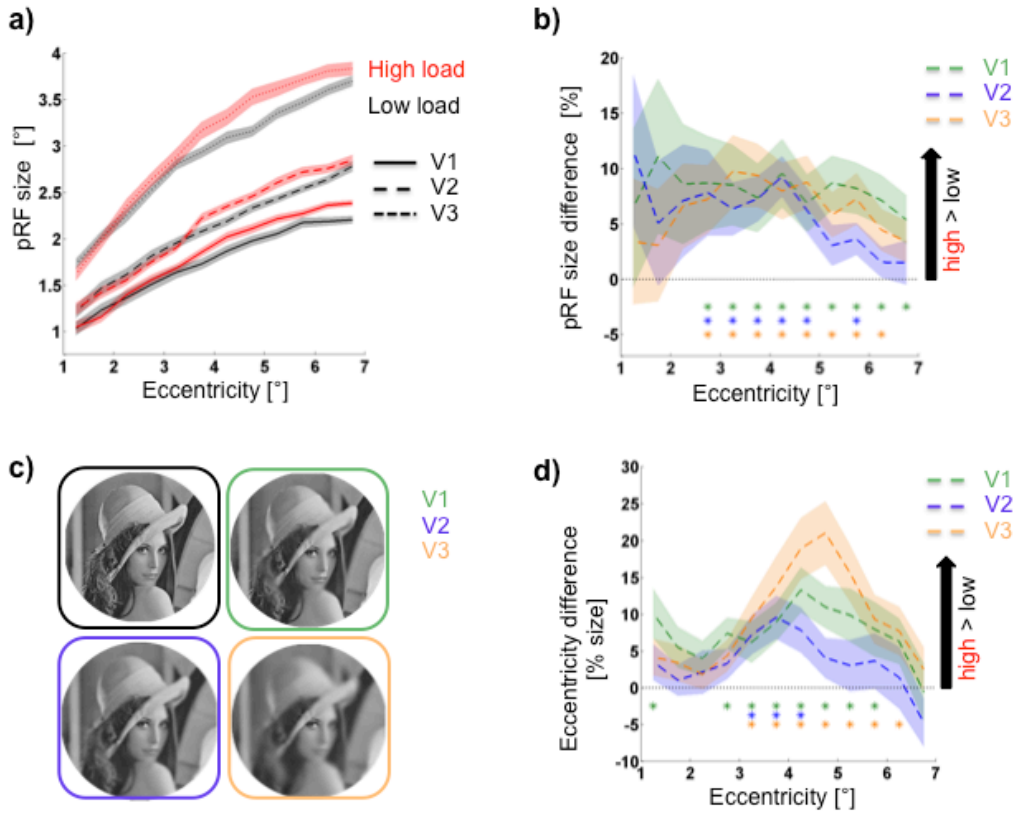

Figure 6-5 Main results. a) pRF size by eccentricity, visual area and condition. pRF sizes increased with eccentricity and along the visual hierarchy. From 3.5 to 6.5 degrees eccentricity perceptual load affected pRF size estimates, with bigger pRFs under high (red) vs. low (black) perceptual load. Lines indicate sample mean with different line styles for each visual area (see inset). Error shades indicate +/- one standard error of the mean (S.E.M.). **b)** Effect of high- vs low perceptual load on pRF size. Relative size differences between conditions were computed and expressed as % size difference. Positive values indicate bigger pRFs under high vs low perceptual load. A difference is clearest from 3-6 degrees eccentricity. Large error for innermost eccentricities reflects instability of the relative measure for very small (absolute) pRFs. Error shades indicate +/- one S.E.M. and asterisks indicate P<.05 for data bins with bin width 0.5 degrees. Color indicates visual area (see inset). **c)** High vs. low perceptual load effects on pRF sizes. The original picture, shown top left, was smoothed in an eccentricity dependent manner according to the absolute changes in pRF size for V1, V2 and V3. The resulting pictures indicate the absolute change in resolution for high vs. low load and are shown in the other three panels (visual area indicated by color; note that this is assuming a picture size extending to 7 degrees eccentricity). **d)** Effect of high- vs low perceptual load on pRF eccentricity. Eccentricity differences between conditions were computed and normalized by pRF size (under low load; expressed as % of pRF size). Positive values indicate an outward pull of pRFs under high vs low perceptual load. Again, the difference is clearest from about 3-6 degrees eccentricity. Error shades indicate +/- one S.E.M. and asterisks indicate P<.05 for data bins (width 0.5 degrees; eccentricity bins according to eccentricity under low load). Color indicates visual area (see inset).

6.3.2.2.2 Effects of swapping hemodynamic response functions between conditions

To further estimate the influence of the HRF on pRF estimates, I re-ran all pRF analyses with swapped HRFs between conditions. Swapping HRFs between conditions had an effect on estimated pRF sizes, with pRFs up to ~3° eccentricity being smaller under high perceptual load, however this effect was not significant. Most importantly, the main results were robust with regard to the HRF used in the model: pRFs from about 3° eccentricity were still significantly bigger and more eccentric under high vs. low perceptual load (**Figure 6-6**).

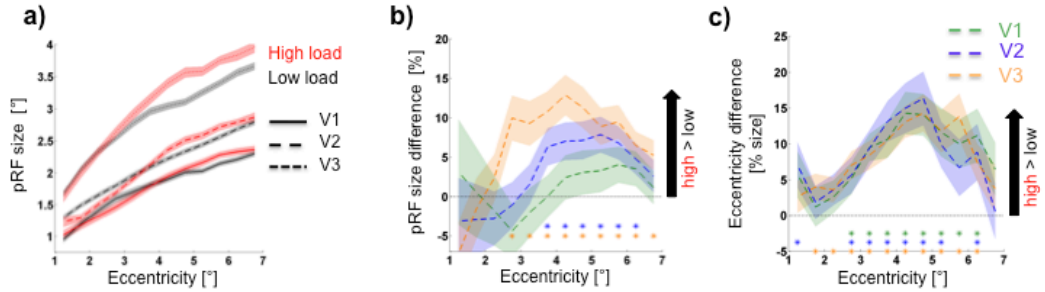

Figure 6-6 Results for re-analysis with swapped HRFs. a) Absolute estimates of pRF size by eccentricity, area and condition. **b)** Relative difference of pRF size between conditions. Positive values indicate larger pRFs under high load. **c**) pRF eccentricity difference between conditions, normalised by pRF size. See legend of **Figure 6-5** for detailed annotations. Swapping HRFs between conditions had an effect on estimated pRF sizes, with pRFs up to ~3° eccentricity being smaller under high perceptual load, however this effect was not significant (**a)** and **b)**). Most importantly, the main results were robust with regard to the HRF used in the model: pRFs from about 3° eccentricity were still significantly bigger (**b)**) and more eccentric (**c)**) under high vs. low perceptual load.

6.4 Discussion

6.4.1 Modulation of pRF size

In this experiment I probed whether perceptual load at fixation can modulate the spatial tuning of neural population responses representing the surrounding visual field from one to seven degrees eccentricity. The first main finding is that the estimated size of surrounding population receptive fields was bigger under high vs. low load. I compared the size of pRFs with center positions from 1-7 degrees eccentricity by calculating the relative size difference between high vs. low load on a vertex-by-vertex basis. Across participants and hemispheres, pRF size significantly increased under high perceptual load for V1, V2 and V3 from ~3-4 degrees eccentricity (**Figure 6-5** a) and b)). In other words the early visual cortex representations of the surrounding visual field were blurred for high perceptual load associated with the central task (illustrated in **Figure 6-5** c).

This resonates with previous findings showing that the withdrawal of *spatial* attention yields perceptual blurring (e.g. (Montagna et al., 2009)), renders spatial information carried by the BOLD signal less precise (Fischer & Whitney, 2009) and increases the size of single neuron receptive fields in area MT (Womelsdorf et al., 2006) (compared to attending a location within the receptive field). Note that here I did not change the locus of attention at all. Participants overtly attended fixation in both conditions and only task-associated perceptual load varied. The finding that perceptual load can reduce the neural spatial resolution for the surrounding visual field points to a previously unknown mechanism of perceptual load, potentially mediating some of its effects. A blurring of distractor representations could

at least contribute to reduced distractor interference (e.g. (Lavie, 1995)), sensitivity (e.g. (Carmel et al., 2011)) and adaptation (e.g. (Rees et al., 1997)).

Participants showed a high degree of gaze stability and thus the observed effect of neural blurring cannot be attributed to eye-movements (see **Figure 6-3**). Another potential confound is the known effects of perceptual load on the amplitude of BOLD responses for the surrounding visual field (e.g. (Schwartz et al., 2005)). Load reduces the BOLD amplitude for surround stimulation, an effect that was replicated in this study via the somewhat lower hemodynamic response to full field photic stimulation under high perceptual load at fixation (**Figure 6-4**). This difference in amplitude was incorporated in the forward model for pRFs by using condition specific HRFs for each participant. Moreover, swapping HRFs between conditions reversed differences in estimated pRF beta parameters but the main changes in spatial preferences were robust to this (**Figure 6-7**; note that beta parameters provide an estimate of pRF amplitude but will also reflect model fit, **Figure 6-8**). Thus I am confident that the observed effects are not mere artifacts of amplitude changes. A final caveat relating amplitude effects is that they have been shown to be eccentricity dependent (Schwartz et al., 2005). Note that I used all voxels in occipital cortex that responded to the full field stimulation (i.e. up to about nine degrees eccentricity) to estimate condition-specific HRFs. A more nuanced approach might estimate condition specific HRFs for eccentricity bands or even single voxels. While the latter would presumably introduce a high degree of noise due to unstable measurements, the former would necessarily run into a

problem of circularity. Estimating pRF eccentricity would already be dependent on some sort of HRF. Finally, the best reason for using condition and hemisphere-specific HRFs (like I did here) is provided by studies explicitly investigating how load effects on BOLD amplitude vary with eccentricity. They show that amplitude effects start to vary with eccentricity from about 8 degrees and thus beyond those I mapped and analyzed here (Figure 8 in (Schwartz et al., 2005)).

It is tempting to speculate that the neural blurring effect for the surround is a trade-off accompanied by *in*creased neural resolution for task-related foveal stimuli. Unfortunately I could not assess the size of foveal population receptive fields. This is where load stimuli were presented and consequently mapping stimuli could not be presented (and mapping of the foveal confluence is difficult in general, c.f. (Schira, Tyler, Breakspear, & Spehar, 2009)). Future experiments might investigate an interaction between covert shifts of (spatial) attention and load. For instance, one could present load tasks at parafoveal locations of each hemifield and instruct people to consecutively solve either of them. At the same time mapping stimuli could traverse the visual field and potentially even cross task locations (e.g. using alpha blended stimuli). It would be interesting to see whether population receptive field containing the attended task location would shrink under high perceptual load.

An interesting difference with regard to single neuron results on the effects of covert spatial attention in hMT+ is that they show the biggest size change of receptive fields near the focus of attention (Anton-Erxleben et al., 2009). In contrast the present population level data from V1-3 reflecting the

effects of perceptual load show the strongest from about 3-4 degrees eccentricity. Note that this eccentricity dependence has to be regarded with caution in the current experiment because measurement errors and thus uncertainty will be relatively bigger for small pRF sizes at inner eccentricities. Still, the reversal of results from single hMT+ neurons and spatial attention is noteworthy. A possible explanation would be that shifts in covert spatial attention (under low perceptual load) only affect the representation of task-associated stimuli (by shifting receptive fields towards those stimuli and shrinking them around the attended location or letting them grow towards it; c.f. (Anton-Erxleben & Carrasco, 2013) and Chapter 1.3.3). In contrast the effects of perceptual load could more directly affect the representation of task-irrelevant stimuli in the surround. Neural blurring could contribute to their suppression as outlined above. Furthermore, neural blurring will effectively shift sensitivity in the surround towards low spatial frequencies. This could not only help to reduce distractor effects. Its adaptive role might also be to help orienting towards stimuli that are big and/or moving (and thus potentially threatening; but c.f. (Forster & Lavie, 2008)). Future experiments could investigate changes in *perceptual* resolution under high vs. low load. This could be done using dual task paradigms, e.g. manipulating load at fixation while testing acuity with Landolt stimuli in the periphery (e.g. (Montagna et al., 2009)) or texture segmentation tasks (e.g. (Yeshurun & Carrasco, 2000)). The latter would be especially interesting because it shows deviating results for exogenous vs. endogenous shifts in spatial attention (c.f. (Yeshurun et al., 2008)). Testing

the effects of perceptual load in such a paradigm might thus allow to directly compare the effect of all three attention manipulations.

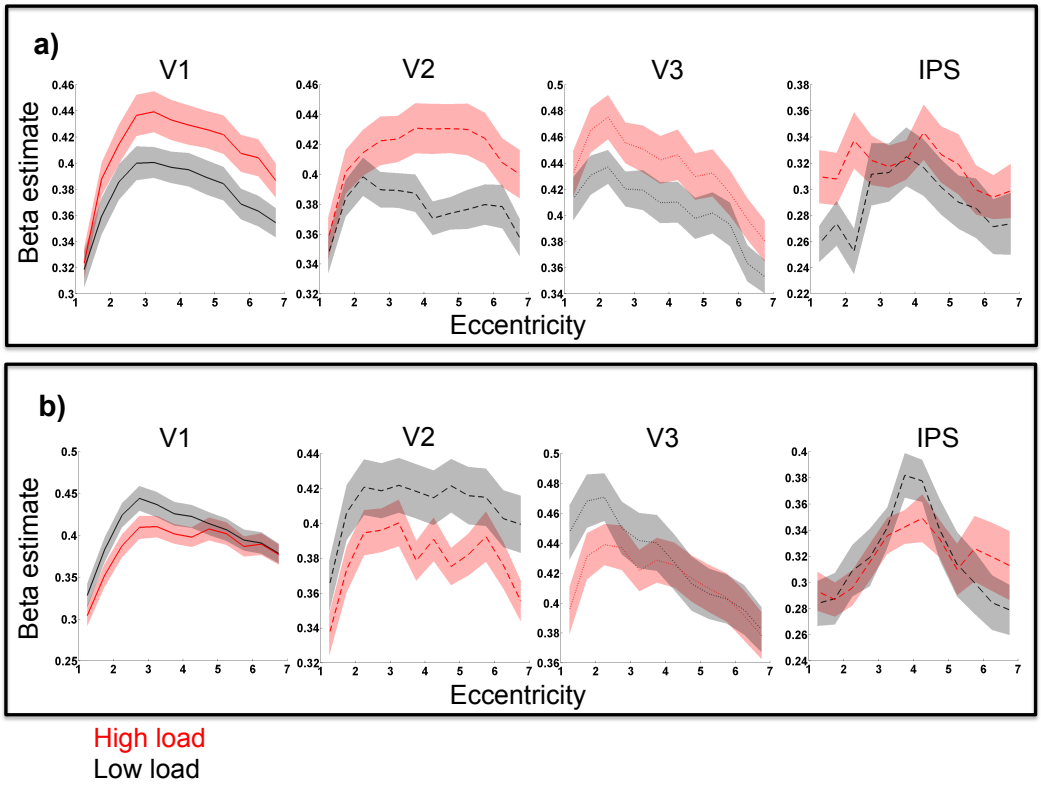

Figure 6-7 Beta estimates for pRFs. a) Beta estimates when using appropriate hemodynamic response functions (HRFs) in the model (i.e. HRFs derived from HRF runs of the corresponding condition). **b)** Beta estimates when using swapped HRFs in the model (i.e. HRFs derived from HRF runs of the other condition). Curves show means calculated for eccentricity bins with width of half a degree visual angle. Error shades indicate +/- one standard error of the mean (S.E.M.). Color indicates condition (black: low load, red: high load). Note that the difference in betas is reversed for swapped HRFs.

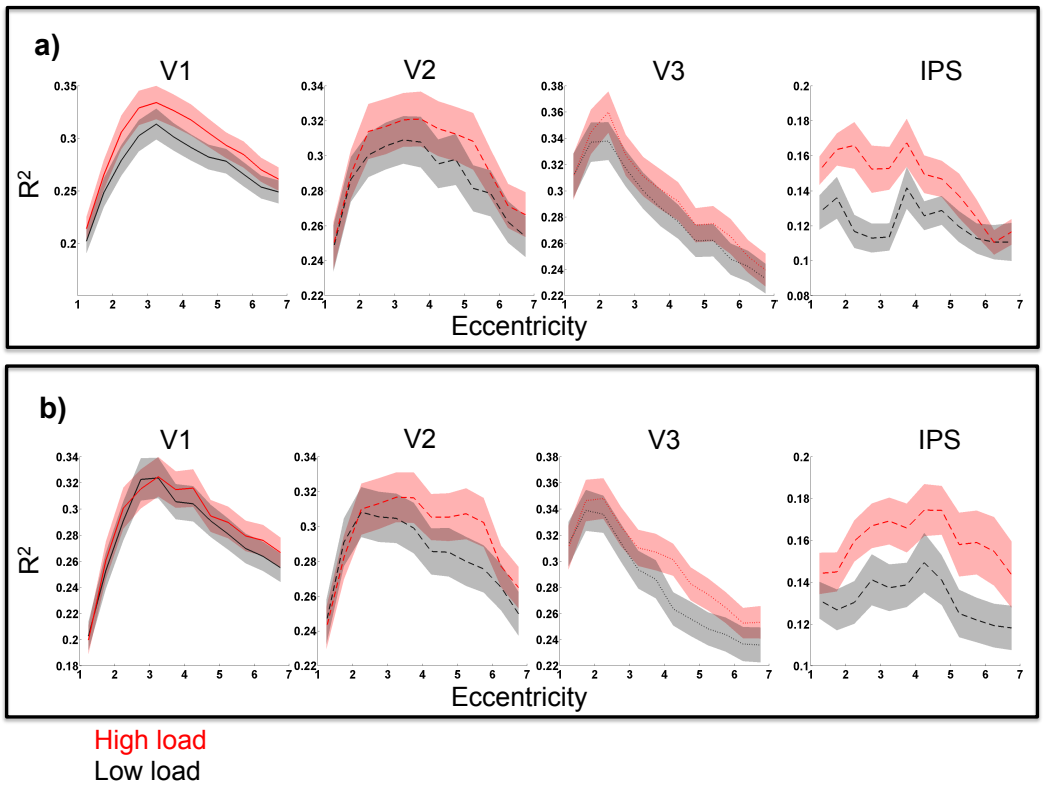

Figure 6-8 Coefficient of determination for pRF fits. a) R^2 of estimates when using appropriate hemodynamic response functions (HRFs) in the model (i.e. HRFs derived from HRF runs of the corresponding condition). **b)** R^2 estimates when using swapped HRFs in the model (i.e. HRFs derived from HRF runs of the other condition). Curves show means calculated for eccentricity bins with a width of half a degree visual angle. Error shades indicate +/- 1 standard error of the mean (S.E.M.). Color indicates condition (black: low load, red: high load).

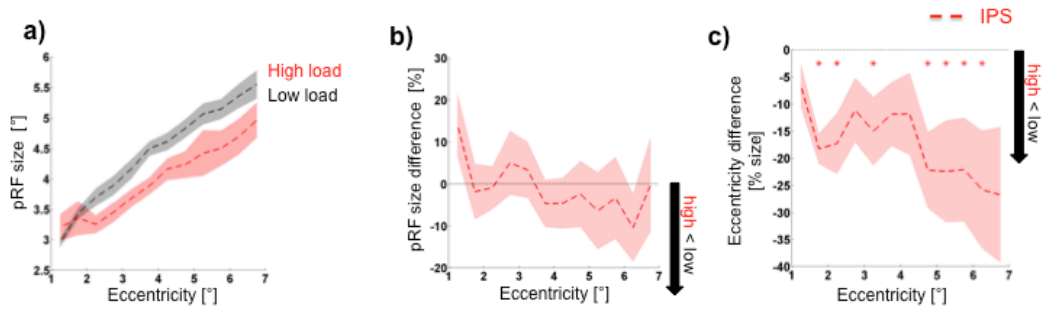

Figure 6-9 Preliminary results for area IPS0/1. a) Absolute estimates of pRF size by eccentricity and condition. **b)** Relative difference of pRF size between conditions. Positive values indicate larger pRFs under high load. **c**) pRF eccentricity difference between conditions, normalized by pRF size. See legend of **Figure 6-5** for detailed annotations Wed did not observe a significant change of pRF size for this area. However, there was a slight, non-significant trend for *de*creased pRF size under high vs. low perceptual load (a), b); mean=-3.44% +/- s.e.m.=4.57%, t_{18}=-0.75, P=0.46). The eccentricity effect on pRFs observed in earlier areas was reversed in IPS0/1 (c)). pRF eccentricity decreased under high perceptual load (mean=-14.03% +/- s.e.m.=4.35%, t16=-3.23, P<0.01). This effect was strongest from about 4.5° eccentricity.

6.4.2 Modulation of pRF location

Apart from pRF size, perceptual load also affected pRF locations. I calculated shifts of pRF center positions (comparing high versus low load) on a vertex by vertex basis. Average center position of pRFs became significantly more eccentric under high load in V1 and V3 with a similar trend in V2. Like the blurring effect, this 'centrifugal' effect on pRFs was strongest from ~3-4 degrees eccentricity. However, it showed a steep decline after a peak at ~4-5 degrees (**Figure 6-5**d)).

This centrifugal effect will shift neural sensitivity for the surrounding visual field away from locations very close to the attended task location. This might serve to help a clearer distinction between task related stimuli

and task-irrelevant distractors and thereby support reduced distractor sensitivity under high perceptual load. Interestingly Lavie proposed already in 1995 (Lavie, 1995) that

'[…], *the load hypothesis may allow a clearer characterization for the role of physical distinctions between relevant and irrelevant stimuli in determining whether processing is selective. The specific suggestion is that such distinctions will be more effective in ensuring the appropriate allocation of attention under situations of high load*.'

It would be interesting to test in behavioral experiments whether the spatial profile of drop-off in distractor interference changes with central perceptual load.

The centrifugal effect contrasts with previous findings that neurons in V4 (Connor et al., 1997), LIP (Ben Hamed et al., 2002) and MT (Womelsdorf et al., 2006) shift their receptive fields *towards* the focus of spatial attention. This could be due to differences between the effects of spatial attention and perceptual load (c.f. above), differences between the behavior of single cells and population responses (some population effects might go undetected in single cell recordings) and/or differences between the neural responses in V1-3 and 'higher' extrastriate areas. My data allowed for a preliminary test of the latter idea. I was able to map IPS0/1 in a subset of 24 hemispheres (the human homologue of LIP). Interestingly, I observed a 'centripetal' change of pRF eccentricity for data from IPS (mean=-14.03%, +/- s.e.m.=4.35%, t_{16}=-3.23, P<0.01; note pRF size in IPS did not change significantly; see Methods section and **Figure 6-9** for more details). This suggests that perceptual load at fixation induces a repulsion effect on the smaller pRFs of early visual cortex, while attracting the bigger pRFs of higher areas. This could make sense if it serves the purpose of better

target-distractor separation. The bigger parafoveal (population) receptive fields of IPS will contain the central task stimuli either way and moving them centripetally means shifting sensitivity towards the location of task-stimuli. The smaller (population) receptive fields of early areas V1-3 however, have separable representations of task stimuli and nearby potential distractor locations. Shifting population receptive fields of the immediate surround outwards might help to reduce distractor interference by separating representations of task and distractor stimuli and by reducing sensitivity for distractors very close to the task stimuli (c.f. above).

Finally, the finding that spatial attention can alter stimulus appearance (e.g. (Fortenbaugh et al., 2011)) has been interpreted to reflect shifted receptive fields and a warping of subjective space due to 'labeled lines' (c.f. (Anton-Erxleben & Carrasco, 2013)). The current results suggest that at least population receptive fields can be shifted in opposing direction between different areas. This raises the question *which* lines would be labeled to anchor subjective space (if any). Combining paradigms probing subjective appearance with perceptual load tasks might further elucidate the relationship between neural changes in spatial preferences and subjective space.

6.4.3 Summary and conclusion

Taken together, the experiment presented in this chapter provided evidence that perceptual load can modulate the size and center positions of pRFs in human V1-3. Population receptive fields in the surrounding visual field are bigger and being pushed outwards under high vs. low central perceptual load. Both of these changes in spatial preference have been

discussed as potential mechanisms supporting distractor suppression and target-distractor separation. I related them to previous findings and proposed further experiments to follow them up.

In particular, experiments combining population receptive field mapping, manipulations of perceptual load and shifts of covert attention would be potentially highly informative. They could investigate whether and what kind of changes in spatial tuning take place at task locations. Furthermore they might be suited to answer the question in what way size and position changes of population receptive fields are linked. The size changes observed for the current results could be a consequence of pRF position changes. For instance, pRF changes might reflect changes in the weighting of neural subpopulations rather than the sum of single unit changes. If this were the case an outward push of pRFs would indicate a greater contribution of more eccentric single unit receptive fields to the voxel's response. This would necessarily entail a pRF size change, as more eccentric receptive fields on average are bigger. Alternatively, changes in pRF size and position might be independent factors (for instance if they reflect changes in single unit receptive fields). Experiments manipulating perceptual load as well as task location relative to fixation could shed light on this question because they would potentially disentangle both factors. If pRF size changes simply follow pRF location changes, pRFs closer to fixation than the task location should shrink under high perceptual load. Another way of gaining more insight into the details of load-induced changes in spatial preferences would be to probe further pRF models (for instance including a suppressive surround; (Zuiderbaan et al., 2012) or model-free

approaches allowing to map asymmetric changes in spatial preferences directly (Lee et al., 2013). Of course learning about the single unit contributions to the population response directly would also be very valuable. Experiments in monkeys or patients, combining electrophysiological cell recordings and fMRI, could provide such information.

Finally, a question falling out of the scope of this chapter is for the neural causes of load-induced changes in spatial preferences in V1-3. These might be mediated in a bottom-up fashion by changes in the spatial preferences of subcortical neural populations, in a top-down manner by parietal or frontal regions (c.f. (Bisley, 2011)). Studies investigating whole brain connectivity might be able to give hints about the interactions between these regions (Haak et al., 2012).

In the next chapter I will investigate a very different question regarding the role of spatial preferences in visual processing. Specifically, whether there is an interaction between spatial and feature preferences in face processing.

7 Does retinotopic stimulus location predict perceptual and neural sensitivity for face features?

7.1 Introduction

Human observers show a stereotypical pattern of gaze behavior towards other people's faces. Typically, the first fixations land on the central upper nose region, just below the eyes (Hsiao & Cottrell, 2008), a landing point that is optimal for rapid face recognition (Peterson & Eckstein, 2012). Subsequent fixations remain restricted to inner face features (e.g. (van Belle et al., 2010)) and observers tend to avoid looking directly at outer features like the chin and upper forehead (cf. Chapter 1.4.5).

This pattern of gaze behavior implies a potential retinotopic bias for the cortical representation of face features. Eyes will tend to consistently appear in the upper visual field, while mouth stimuli will predominantly appear in the lower visual field. This is because eyes will only appear in the lower visual field when fixating the upper forehead or above and mouth stimuli will only appear in the upper visual field when fixating the chin or below (neither of which observers typically do). But does retinotopic location matter in the context of face perception?

Some models of object and face recognition propose a hierarchy of processing along the ventral visual stream from view dependent to more abstract representations. Representations at the later stages are thought to be closest to the outcome of object recognition and largely invariant to low-

level properties like stimulus location (e.g. (DiCarlo & Cox, 2007; Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007)). This view is supported by the tuning properties of many neurons in inferior temporal cortex (IT) of rhesus monkeys, which have a preference for shapes, color or specific object categories (like faces) but large receptive fields covering most of the central visual field (e.g. (Desimone et al., 1984; Gross et al., 1972; Gross, 1992; Ito, Tamura, Fujita, & Tanaka, 1995; Tovee et al., 1994)). Additional support for location invariance comes from priming studies in humans that found priming advantages to spread across the visual field (Biederman, Cooper, Kourtzi, Sinha, & Wagemans, 2009; Biederman & Cooper, 1991; Ellis, Allport, Humphreys, & Collis, 1989).

However, recently the notion of location invariant object and face processing has been challenged (Kravitz et al., 2013; Kravitz, Vinson, & Baker, 2008). In humans, perceptual experiments found face adaptation effects to be retinotopically confined (Afraz & Cavanagh, 2009, 2008; Dickinson, Mighall, Almeida, Bell, & Badcock, 2012; Zimmer & Kovács, 2011). Furthermore, (Afraz et al., 2010) found idiosyncratic but reliable retinotopic heterogeneity for age and gender judgments based on faces. Human fMRI studies investigating BOLD signals in human face sensitive cortex found evidence for a role of stimulus location with regard to the overall response amplitude in face sensitive cortex (e.g. (Hemond, Kanwisher, & Op de Beeck, 2007; Schwarzlose, Swisher, Dang, & Kanwisher, 2008; Yue, Cassidy, Devaney, Holt, & Tootell, 2011)), as well as to the patterns of activation within those areas (Chan, Kravitz, Truong, Arizpe, & Baker, 2010; Schwarzlose et al., 2008) and to response adaptation ((Kovács,

Cziraki, Vidnyánszky, Schweinberger, & Greenlee, 2008); see Discussion section for more details). Finally, single neuron studies in monkey IT found evidence for a range of receptive field sizes in IT and some scatter of receptive fields across the visual field (DiCarlo & Maunsell, 2003; Logothetis, Pauls, & Poggio, 1995; Op De Beeck & Vogels, 2000).

Taken together, face features differ in the frequency with which they will appear in different parts of the visual field and there is evidence to suggest a role of stimulus location in face perception. This leads to the question whether spatial and feature preferences are linked. A tuning to canonical visual field positions has recently been shown for face halves and upper limbs with regard to the visual hemi-field in which they appear (Chan et al., 2010). So, for example, there is a recognition advantage for stimuli representing the right hand or the right halves of faces when presented in the left visual hemi-field and vice versa.

Here, I asked whether observers showed a similar kind of tuning to canonical retinotopic locations for isolated face features. Specifically, I hypothesized that recognition of eye and mouth stimuli would be better when presented at *canonical* versus *swapped* visual field locations. This hypothesis was tested in a two-alternative forced choice (2AFC) task comparing recognition performance for either type of face feature across different retinotopic locations.

7.2 Behavioral Experiment

7.2.1 Method

7.2.1.1 Participants

18 healthy participants from the University College London (UCL) participant pool took part in the behavioral study (aged 19 to 52 years, mean: 27 yrs, SD: 9 yrs; 11 females; 3 left-handed). All participants had normal or corrected to normal vision. Written informed consent was obtained from each participant and the study was approved by the UCL ethics committee.

7.2.1.2 Stimuli

The face feature stimuli stem from a set of 92 frontal photographs of faces with neutral expression (stimuli were kindly provided by Dr. Linda Henriksson and Dr. Nikolaus Kriegeskorte, Cognition and Brain Sciences Unit, Cambridge University). 32 of these images were excluded because they contained easily identifiable eye or mouth features (e.g. facial hair), leaving images from 60 faces in the dataset, 21 of which were male. This dataset was used to form 27 face candidate pairs. Faces were categorized according to three binary variables encoding gender, skin and eye color (dark/light). Then an algorithm was used to pseudorandomly choose faces, forming pairs of images that were matched with regard to all three variables.

Each of the 54 candidate faces yielded a set of three candidate face features (left eye, right eye and mouth images, respectively). These face feature images were sampled according to a grid that was overlaid on each face (**Figure 7-1**). Feature images were square and presented at a visual

angle of 3.5 degrees (viewing distance 80 cm). The outer edge of the image was overlaid with a grey fringe that was ~0.4 degrees wide and softened the edge between image and background. Participants saw an 8 bit grey-scaled version of the feature images that was displayed on a gray background and with a dynamic noise mask overlay (see below). Stimuli were shown on a liquid crystal display monitor (LCD; Samsung SyncMaster 2233RZ) with a refresh rate of 120 Hz and a spatial resolution of 1680 by 1050 pixels.

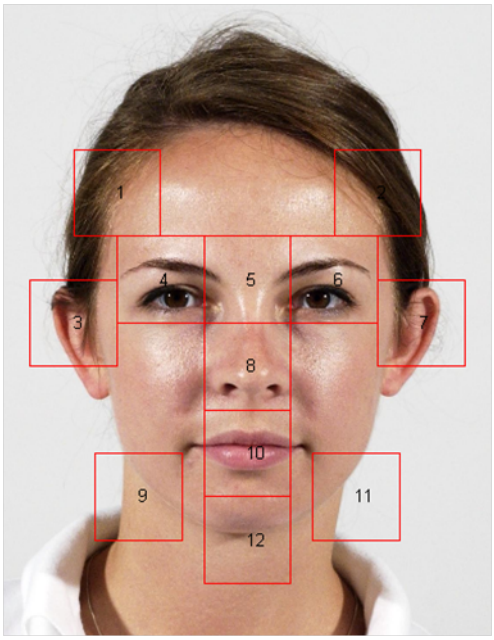

Figure 7-1 Sampling grid for face features. A symmetric grid of squares was overlaid the face images, such that two squares were centered on the left and right eye region as shown in the example. This way a further square tile was centered on the mouth region. The respective tile regions were cut out and served as feature images in the experiment. Image kindly provided by Dr. Linda Henriksson and Dr. Nikolaus Kriegeskorte, Cognition and Brain Sciences Unit, MRC, Cambridge.

7.2.1.3 Procedure

Participants sat on a chair with their head in a chin rest at a viewing distance of 80 cm. Each trial began with the presentation of a blue fixation dot at the middle of the screen and on grey background (0.1 degrees visual angle in diameter). After 500 ms the image of a face feature flashed up for 200 ms, overlaid by a noise mask that lasted until 450 ms post image onset. Each frame of the noise mask consisted of a random intensity image of the same size as the face feature images and with pixel intensities drawn from a uniform distribution ranging from 0 to 128 (corresponding to the background gray level; the luminance range of the display was not linearized). Pixel intensities of overlay frames corresponded to the numerical average of the noise mask and face feature images.

Immediately after the offset of the noise mask, the fixation dot turned green and two candidate images appeared on the screen, prompting the participant to indicate which of them flashed up earlier. Candidate images were scaled versions (75% original size) of the target image and the image of the corresponding face feature in the matched candidate face (see above). During presentation of the candidate images the fixation dot was not presented. Candidate images were centered at the height at which the fixation dot was shown previously and centered at 2 degrees visual angle left and right of this location. Participants used the arrow keys of a standard keyboard to shift a blue selection rectangle surrounding either candidate image and confirmed their choice with the space bar. The selection rectangle then turned either green or red for 300 ms, indicating the answer was correct or incorrect, respectively.

Each participant completed nine blocks of 48 trials each. A block contained trials corresponding to three sets of candidate images or six faces. Feature images of each face served as stimuli in a total of eight trials, corresponding to eight trial types. Trial types were defined by face feature (left eye, right eye, mouth) and stimulus position. There were three main stimulus positions that corresponded to the locations of the left eye, right eye and mouth segments in the original image, assuming fixation slightly above the nose. Specifically, the mouth position was centered on the vertical meridian at 4.4 degrees visual angle below fixation and the left and right eye positions were centered 2.6 degrees above fixation and shifted 3.5 degrees to the left or right, respectively. Note that all center positions had equal eccentricity (4.4 degrees).

The eight trial types corresponded to either eye image appearing at its 'correct' location (two trial types corresponding to an *upper visual field eye* condition) or the mouth position (two trial types corresponding to a *lower visual field eye* condition) and the mouth image appearing at either eye position (two trial types corresponding to an *upper visual field mouth* condition) or the mouth image appearing at the mouth position. To balance the design, the trials of the latter trial type were repeated in each block, yielding two (dummy coded) trial types corresponding to a *lower visual field mouth* condition. The order of trials was randomized within each block and in each trial the exact stimulus position was determined as the main location (corresponding to trial type) plus a random offset to avoid adaptation or fatigue. Spatial scatter on the horizontal and vertical plane were drawn independently from a Gaussian distribution centered on zero with a

standard deviation of 0.35 degrees visual angle and clipped at two standard deviations.

To ensure fixation, gaze direction was monitored with an infrared eyetracker (Cambridge Research Systems) operating monocularly at 200 Hz. Gaze data were collected for 13 participants (calibration of the eyetracker or recording failed for the remaining five participants). For these 13 participants gaze direction could successfully be tracked for an average of 87.11% of trials (s.e.m.=5.29%).

7.2.1.4 Analysis

All statistical analyses of behavioral data were performed in MATLAB (Mathworks, Ltd.) and PASW 18.0 (SPSS inc./IBM). To test for an interaction between face feature and visual field position, the proportion of correct answers was averaged for each participant and condition (i.e. *location* (upper/lower visual field) by *stimulus* (eye/mouth)). The resulting values were compared across conditions using a repeated measures general linear model and post-hoc *t*-tests.

An indicator of general fixation compliance was computed as the median absolute deviation of gaze direction during stimulus presentation (excluding the post-stimulus noise mask). This was done separately for the vertical and horizontal axes. Additionally, an indicator of gaze bias towards the stimulus was computed. For this, gaze direction on the vertical axis was compared between trials in which stimuli were presented in the upper vs. lower visual field. A bias index was defined as the median difference in vertical eye position between these trial types. This bias index was also

calculated separately for and compared between trials with *canonical* vs. *swapped* stimulus locations. Finally, to test whether lack of fixation compliance predicted the hypothesized effect, a correlation between individual effect size and propensity for eye movements was computed across participants. For this, effect size was defined as the individual recognition advantage in *canonical* vs. *swapped* trials and the propensity for eye movements was defined as the median absolute deviation of gaze direction during stimulus presentation (averaged across the horizontal and vertical axes).

7.2.2 Results

7.2.2.1 Recognition performance

There was a significant recognition advantage for eye over mouth stimuli ($F_{1,17}$=4.95, P<.05), but no main effect of stimulus position ($F_{1,17}$=0.43, P=.52; cf. **Table 7-1** for means and standard errors). Crucially, there was a highly significant interaction between face feature and visual field location ($F_{1,17}$=21.87, P<.001). Specifically, post-hoc t-tests revealed that recognition of eyes was significantly better in the upper than lower visual field (t_{17}=3.34, P<.01), while the reverse was true for recognition of mouth stimuli (t_{17}=-3.40, P<.01; **Figure 7-2**; cf. **Table 7-1** for means and standard errors of recognition performance).

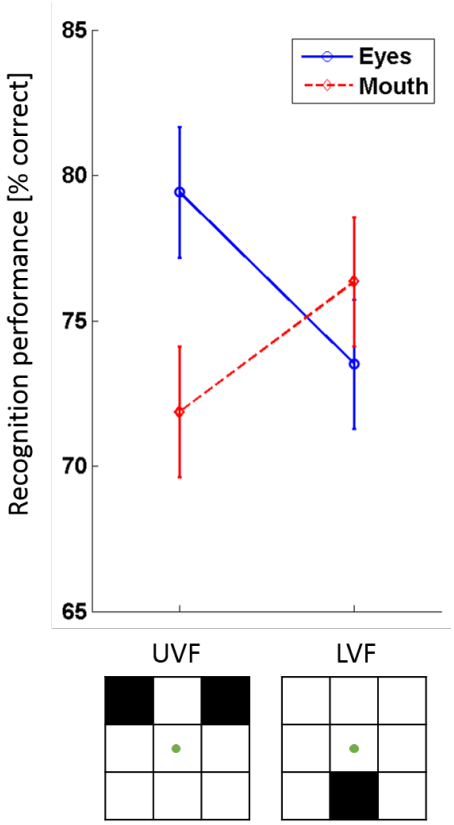

Figure 7-2 Stimulus by location interaction for recognition of facial features. Recognition performance for eye stimuli is plotted in blue, recognition for mouth stimuli in red. The left and right hand sides of the plot indicate recognition performance for upper and lower visual field locations, respectively. The tile schemas below the x-axis show the respective visual field locations as filled areas (green dot indicates fixation location; see Procedure for details). Error bars indicate mean recognition performance +/- one standard error of the mean.

Table 7-1 Recognition performance for face features by visual field location. Cells contain the mean and standard error of the mean (S.E.M.) of % correct answers across participants. See Stimuli and Procedure for details.

	Upper visual field	**Lower visual field**	
Eye Stimuli	79.49% (2.25%)	73.51% (2.23%)	76.47% (2.06%)
Mouth stimuli	71.86% (1.77%)	76.34% (1.62%)	74.10% (1.56%)
	75.64% (1.90%)	74.92% (1.76%)	

7.2.2.2 Gaze data

General fixation compliance was good, with an average median absolute deviation from fixation of 0.48 (+/-0.05 S.E.M.) and 0.68 (+/- 0.12 S.E.M.) degrees visual angle on the horizontal and vertical axes, respectively (**Figure 7-3** *a*)). Furthermore, deviation from fixation was not significantly biased towards stimuli in either *canonical* (average bias: 0.11 (+/- 0.09 S.E.M.) degrees visual angle; t_{12}=1.15, *P*=0.27) or *swapped* trials (average bias: 0.13 (+/- 0.10 S.E.M.) degrees visual angle; t_{12}=1.36, *P*=0.20) and the difference in bias between these trial types was not significant either (t_{12}=-1.27, *P*=0.23; cf. **Figure 7-3** *b*)). Finally, individual differences in the propensity for eye movements did not predict individual differences in the size of the hypothesized effect (r_{11}=-.24, *P*=.44; **Figure 7-3** *c*)).

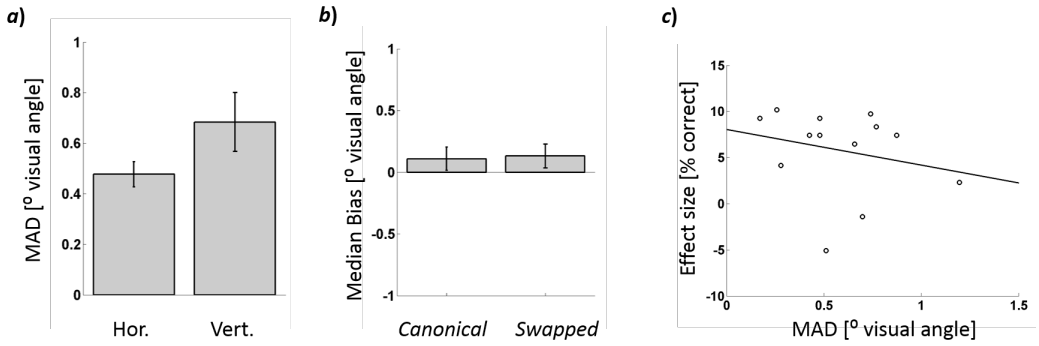

Figure 7-3 Gaze behavior. *a*) Median absolute deviation from fixation during stimulus presentation (in degrees visual angle). The left and right hand bars indicate deviation along the horizontal and vertical axes, respectively. ***b*)** Median bias of gaze direction towards stimuli (along the vertical axis and in degrees visual angle). The left and right hand bars indicate bias during presentation of stimuli in canonical and swapped locations, respectively. Canonical locations are in the upper visual field for eye stimuli and in the lower visual field for mouth stimuli. ***c*)** Scatter plot of individual effect size vs. individual deviation from fixation. Effect size is defined as the recognition advantage for stimuli presented in canonical vs. swapped locations (in % correct answers). Deviation from fixation is the median absolute deviation in degrees visual angle (averaged across the horizontal and vertical axes). Each circle represents data from one participant. A line indicates the least squares fit to the data. All error bars indicate the mean across participants +/- one S.E.M.. See Methods and Results sections for more details.

7.2.3 Discussion

Participants showed a clear recognition advantage for face features presented at canonical retinotopic locations (assuming fixation on the central upper nose (Hsiao & Cottrell, 2008)). Additional eye tracking data confirmed fixation compliance. Gaze direction was not biased towards stimuli in either condition.

This pattern of results is in line with the hypothesis that perception of face features is tuned to input statistics. The processing of eye and mouth

stimuli appears more efficient at canonical visual field locations. This echoes the results of (Chan et al., 2010) who found a similar recognition advantage for face halves and right and left limbs presented in the canonical visual hemi-field. The current result also allows for hypotheses regarding the *neural* processing of face features.

As mentioned above, face sensitive neurons in macaque IT on average have large receptive fields that tend to overlap with the fovea (e.g. (Tovee et al., 1994)). However, a number of studies systematically investigating the spatial tuning of IT neurons found considerable variance of receptive field sizes as well as scatter of receptive fields across the central visual field (DiCarlo & Maunsell, 2003; Logothetis et al., 1995; Op De Beeck & Vogels, 2000). Similarly, fMRI responses in human OFA and FFA are not entirely location invariant. Response amplitudes in these areas are slightly higher for contralateral stimuli (Hemond et al., 2007) and decrease with stimulus eccentricity (in a way reflecting the cortical magnification of V1; (Schwarzlose et al., 2008; Yue et al., 2011)). Stimulus adaptation of fMRI responses in right OFA is only found if adaptor and test faces are shown in the contralateral hemifield, while right FFA adapts to stimuli in either hemifield and adaptation generalizes across hemifields for FFA (Kovács et al., 2008). Furthermore, response patterns in OFA and FFA carry location invariant information about stimulus category but *also* category invariant information about stimulus location (Schwarzlose et al., 2008). In summary, face stimulus location plays a significant role in face sensitive areas of cortex.

Furthermore, although many studies investigated neural responses to whole faces there are some exceptions using face parts as stimuli. (Perrett et al., 1982) investigated response properties of face sensitive neurons in macaque superior temporal sulcus (STS). More than 70% of neurons in this area responded to isolated face features (like an eye or a mouth) as well as whole faces. The stimulus tuning profile of cells was varied. Some neurons only responded in the presence of one specific feature while others appeared to have a broader stimulus tuning function, echoing the variance in spatial tuning among IT neurons (c.f. their figures 9 and 10). Face sensitive neurons in STS also show tuning to the arrangement of specific face parts (like the inter-eye distance or size of eyes; (Freiwald et al., 2009)). Crucially, a recent study found both spatial and feature tuning in the same neurons of a posterior face sensitive patch in monkey IT (investigating only the left hemisphere; (Issa & DiCarlo, 2012)). Most neurons in this area showed a strong preference for eye stimuli, regardless of whether they were presented in the context of a face or not (and especially for early response components). Interestingly, the same neurons showed a retinotopic preference for the contralateral upper visual field, both when mapped with whole faces and face parts.

In humans, fMRI responses in OFA and FFA depend on the presence of face parts and not necessarily their arrangement as a face (although FFA responses to scrambled faces are slightly diminished; (Liu, Harris, & Kanwisher, 2010)). OFA seems particularly responsive to parts of a face, like the inner features or face outline (Nichols et al., 2010). TMS over right OFA disrupts face recognition when it is based on feature discrimination but not

when it is based on discrimination of spatial arrangements (Pitcher et al., 2007). Data from a preliminary study suggest that OFA and FFA respond to isolated face features, with a possible map-like organization of responses in OFA (Kriegeskorte, Mur, & Henriksson, 2013). Furthermore, the typical N170 response to faces can be driven by the isolated contra-lateral eye region ((Smith, Gosselin, & Schyns, 2004); c.f. (Smith, Fries, Gosselin, Goebel, & Schyns, 2009)). Finally, the above mentioned study by (Chan et al., 2010) found that the recognition advantage for face halves in the canonical visual hemi-field was reflected in stimulus information carried by evoked patterns of activation in right FFA. Patterns evoked by different stimulus categories (face halves vs. limbs) were significantly better separable if stimuli were shown in the canonical hemi-field (this was also true for different categories of limb stimuli and patterns in the right extrastriate body area, EBA).

Taken together, there is converging evidence for some spatial tuning and tuning to specific features in face sensitive areas of cortex. At least one electrophysiological study in monkeys (Issa & DiCarlo, 2012) and one fMRI study in humans (Chan et al., 2010) suggest that spatial and feature tuning of neural responses can systematically co-vary.

Such a correlation between spatial and feature preferences could explain the results of the behavioral experiment presented here. Recognition of eye and mouth stimuli might rely on responses of neural populations tuned to the respective face feature. This would predict the observed patterns of results if neural populations preferring eye and mouth stimuli also had a spatial preference for the upper and lower visual field, respectively.

How could this hypothesis be tested in healthy human participants? Tuning properties of neurons can be reflected in subtle preferences of voxels – especially if neurons cluster according to their tuning properties (as suggested by e.g. the results of (Issa & DiCarlo, 2012)). These voxel biases in turn can shape the overall pattern of activation in a given area for a given type of stimulus,. This in turn enables decoding algorithms to predict stimulus category based on patterns of activation. A pre-condition for testing the proposed hypothesis is that eye and mouth stimuli *will* evoke separable patterns of activation in face sensitive patches of cortex. This would indicate responses to eye and mouth stimuli are dominated somewhat by voxels with a slight preference for the respective type of stimulus.

The null hypothesis assumes no correlation between spatial and feature tunings of neural populations. Any spatial preferences of voxels biased towards eye and mouth stimuli would be balanced and averaged out across voxels. Thus, overall responses should be location invariant (or at least showing identical spatial preferences for either stimulus type). Therefore the null hypothesis predicts that the separability of patterns will be invariant to stimulus location (**Figure 7-4**, left hand side).

The alternative hypothesis proposes that feature preferences are associated with a spatial preference for canonical stimulus locations. So, voxels dominated by eye and mouth preferring neurons or neuronal populations would also have a preference for the upper and lower visual field, respectively (**Figure 7-4**, right hand side). Stimuli shown at canonical locations will evoke stronger responses in voxels preferring the respective

feature because they also coincide with their spatial preference. At the same time they will evoke less response in voxels with a preference for the other stimulus category because they also are presented at a suboptimal location for these voxels. Consequentially the alternative hypothesis predicts that pattern separability in face sensitive cortices will be better for stimuli presented at *canonical* versus *swapped* locations.

7.3 fMRI experiment

7.3.1 Method

7.3.1.1 Participants

21 healthy participants from the University College London (UCL) participant pool took part in the fMRI study (aged 20 to 32 years, mean: 25yrs, SD: 4yrs; 14 females; 1 left-handed). All participants had normal or corrected to normal vision. Written informed consent was obtained from each participant and the study was approved by the UCL ethics committee.

7.3.1.2 Stimuli

7.3.1.2.1 Main experiment

Stimuli were identical to the behavioral experiment (see above). Participants viewed stimuli via a mirror mounted at the head coil at a viewing distance of ~73 cm, resulting in an effective stimulus size of 3.2 x 3.2 degree visual angle.

7.3.1.2.2 Face localizer

Each participant completed an additional localizer run consisting of alternating blocks of images depicting faces and everyday objects. Face images were frontal color photographs of 10 male and 10 female faces with neutral expression taken from the PUT face database ((Kasiński, Florek, & Schmidt, 2008); https://biometrics.cie.put.poznan.pl/). Object images were color photographs depicting 20 everyday objects (e.g. a boot, a wall clock or a flower). Jonas Kubilius and Dr. Lee de-Witt, KU Leuven, kindly provided these images. All localizer images were quadratic, shown on a grey background at the center of the screen and at a size of 4 x 4 degrees visual angle. The center of the images contained a circular cut out (width: 0.4 degrees visual angle) with a black fixation dot (width: 0.1 degrees visual angle).

All stimuli were presented with a projector at a resolution of 1024x1280 pixels and with a refresh rate of 60 Hz. Stimulus presentation was controlled using MATLAB (Mathworks, Ltd.) and the Psychophysics Toolbox 3 extension (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) (http://psychtoolbox.org).

7.3.1.3 Procedure

7.3.1.3.1 Main experiment

Each participant completed 10 scanning runs that lasted just under four minutes each. To motivate fixation participants were instructed to ignore the face stimuli flashing in the periphery and performed a dimming task at fixation throughout. At the beginning of each run a black fixation dot

(0.1 degrees visual angle in diameter) was presented at the center of the screen on a grey background. Every 750 ms the fixation dot changed its color for 500 ms from black to one of 13 colors (three shades of blue, green, yellow, orange, pink, three shades of purple, gray, white, red). Participants were asked to press a button if and only if the dot color changed to red.

In parallel to the ongoing fixation task participants were presented with a face feature stimulus every four seconds. This amounted to 54 stimulus trials per run, 6 of which were baseline trials during which no stimulus was presented (apart from the ongoing fixation task). The remaining 48 stimulus trials were very similar to the behavioral experiment, except for participants performing the fixation task instead of a recognition task.

Each run contained stimuli from two face photographs (i.e. two mouth stimuli and two pairs of eye stimuli) that randomly changed every two runs. Face feature stimuli (left eye/right eye/left mouth/right mouth with dummy coding for the latter) and location (respective eye/mouth) were combined to result in eight trial types identical to the behavioral experiment and the corresponding four conditions (upper/lower visual field x eye/mouth stimuli). Each of the eight trial types was repeated three times for stimuli from either face photograph resulting in a total of 48 trials (plus six baseline trials) per run. The order of trials was pseudo-randomized, such that repetitions did not immediately follow each other.

At the beginning of each trial the stimulus image was briefly presented for 200 ms, overlaid by a dynamic noise mask that lasted until 450 ms after the stimulus image onset, just as in the behavioral experiment.

The mouth location was centered below fixation, on the vertical meridian at an eccentricity of 4 degrees visual angle. The left and right eye positions were centered 2.4 degrees above fixation and shifted laterally by 3.2 degrees, resulting in 4 degrees eccentricity. As in the behavioral experiment, the exact stimulus location in each trial varied randomly with the scatter on the horizontal and vertical axes drawn from a Gaussian distribution (SD: 0.16 degrees visual angle; maximum scatter 2 SD).

7.3.1.3.2 Face localizer

Face localizer runs consisted of 20 alternating blocks of face and object images. Each block lasted just over 16 seconds (9 volumes at a TR of 1.785 s) and contained the 20 images of the respective set in randomized order, resulting in a run duration of just under five and a half minutes. Images were presented briefly for 400ms with an inter-stimulus interval of 400ms, while the black fixation dot was shown throughout.

Participants were instructed to perform a 1-back task. In each block a randomly determined image of the sequence was replaced by the preceding image of the sequence, resulting in a 1-back target. Participants were instructed to press a button whenever an image was identical to the immediately preceding one.

7.3.1.3.3 Retinotopic mapping

Apart from the main experiment and face localizer runs, each participant also completed two runs of standard retinotopic mapping (each

lasting under five minutes and using a flashing wedge and ring type of stimulus). These data were not used for the present experiment.

7.3.1.4 Image acquisition and pre-processing

All functional and structural scans were obtained with a Tim Trio 3T scanner (Siemens Medical Systems, Erlangen, Germany), using a 32-channel head coil. However, the front part of the head coil was removed for functional scans, leaving 20 effective channels (this way restrictions of participants' field of view were minimized). Functional images for the main experiment were acquired with a gradient echo planar imaging (EPI) sequence (2.3 mm isotropic resolution, matrix size 96 x 96, 21 transverse slices per volume, acquired in interleaved order and centered on the occipital and inferior temporal cortex; slice acquisition time 85 ms, TE 37 ms, TR 1.785 s). I acquired 129 volumes for each run of the main experiment and 185 volumes per localizer run. After five runs of the main experiment B0 field maps were acquired to correct for geometric distortions in the functional images caused by heterogeneities in the B0 magnetic field (double-echo FLASH sequence with a short TE of 10 ms and a long TE of 12.46 ms, 3x3x2 mm, 1 mm gap). Finally, two T1-weighted structural images were acquired for each participant. The first structural image was obtained with the front part of the head coil removed, using an MPRAGE sequence (1 mm isotropic resolution, 176 sagittal slices, matrix size 256 x 215, TE 2.97 ms, TR 1900 ms). For the second structural image the full 32-channel head coil was used with a 3D MDEFT sequence (Deichmann et al., 2004); 1 mm

isotropic resolution, 176 sagittal partitions, matrix size 256 x 240, TE 2.48 ms, TR 7.92 ms, TI 910 ms).

All image files were converted to NIfTI format and pre-processed using SPM 8 (http://www.fil.ion.ucl.ac.uk/spm/software/spm8/). The first five volumes for each run were discarded to allow for the T1 signal to reach steady state. The remaining functional images were mean bias corrected, realigned, unwarped (using voxel displacement maps generated from the fieldmaps (Hutton et al., 2002)), co-registered (with the respective anatomical MDEFT scan for each participant, using the MPRAGE scan as an intermediate step) and smoothed with a 5 mm Gaussian kernel. Images were pre-processed for each run independently. The anatomical MDEFT scan was used to reconstruct the cortical surface with FreeSurfer (http://surfer.nmr.mgh.harvard.edu).

7.3.1.5 Data analysis

7.3.1.5.1 Regions of interest

To define regions of interest (ROIs) a general linear model was specified for the localizer run of each participant. The model contained one regressor for blocks of face stimuli and one regressor for blocks of object stimuli (boxcar regressors convolved with a canonical hemodynamic response function). Additional regressors of no interest were modelled for the six motion parameters estimated during re-alignment. The parameters of the model were estimated and a contrast images and t-map calculated for the parameter contrast *face – object stimuli* for each participant.

The resulting t-maps were thresholded at $t>2$ and projected onto the inflated cortical surface for each participant in FreeSurfer. Then contiguous patches of activation were identified and delineated at anatomical locations corresponding to those described by (Weiner & Grill-Spector, 2013): mid fusiform face area (mFus), posterior fusiform face area (pFus) and inferior occipital gyrus face area (IOG)).

Because activations in IOG tended to be diffuse and were not detectable in 12 hemispheres, the IOG region of interest was defined anatomically on an individual basis. For this the FreeSurfer parcellation algorithm (Destrieux et al., 2010) was used to label the inferior occipital gyrus and sulcus on each hemisphere.

Fusiform patches of activation could be detected in each of the 42 hemispheres. However, in 24 hemispheres there was a single patch of fusiform activation that could not be separated into mFus and pFus. Therefore fusiform patches of face sensitive cortex were assigned a single FFA label for each hemisphere, resulting in a total of four ROIs per participant (left and right IOG as well as left and right FFA; c.f. **Figure 7-5**).

Finally, all region of interest labels were transformed into volume space and used as binary masks for the analyses described below.

7.3.1.5.2 Multivoxel pattern analysis

Separate general linear models were run for each run and each participant. Each general linear model contained regressors for each of the 8 trial types plus one regressor for baseline trials (boxcar regressors convolved with a canonical hemodynamic response function). Additional

regressors of no interest were modelled for the six motion parameters estimated during re-alignment. The general linear models for each run and each participant were estimated and contrast images for each of the 8 trial types (per run) calculated. This resulted in separate contrast images and t-maps for each trial type, run and participant. These t-maps were masked with the regions of interest (see above) and the resulting patterns were vectorised (i.e. collapsed into a single column of data with entries corresponding to voxels in the original data space).

The aim of the decoding analysis was to decode stimulus identity from activation patterns (i.e. whether an eye or mouth stimulus was presented in a given trial) and to compare the accuracies of decoders across conditions (i.e. whether decoding accuracy varied for *canonical* and *swapped* stimulus locations, c.f. **Figure 7-4**). Stimulus decoding was performed using custom code and the linear support vector machine (lSVM) implemented in the Bioinformatics toolbox for MATLAB (version R2013a, http://www.mathworks.com). Data from each condition were used for training and testing of separate classifiers to get condition-specific decoding accuracies. To avoid assigning a single classification label to stimuli shown at different locations, data within each condition were further subdivided according to visual hemi-field. That is, data from left eye and left mouth stimuli were decoded separately as well as data from right eye and right mouth stimuli. For the *canonical* condition left and right mouth stimuli were dummy-coded (see above). Classification accuracies did not significantly differ between visual hemifields in either condition or any ROI and were thus averaged across visual hemi-fields for each condition.

Classifiers were trained and tested for accuracy in a jackknife procedure. In each iteration the (condition and hemi-field specific) data from all runs but one served as training data and the (condition and hemi-field specific) data from the remaining run was used to test the prediction accuracy of the lSVM. Accuracies were stored and averaged across iterations at the end of this procedure, and the whole procedure was applied to each region of interest independently, yielding a single classification accuracy for each condition and ROI. Statistical analysis of the resulting accuracies was performed in MATLAB. Accuracies were compared against chance level by subtracting 0.5 and using one sample t-tests. Accuracies were compared between conditions using paired t-tests.

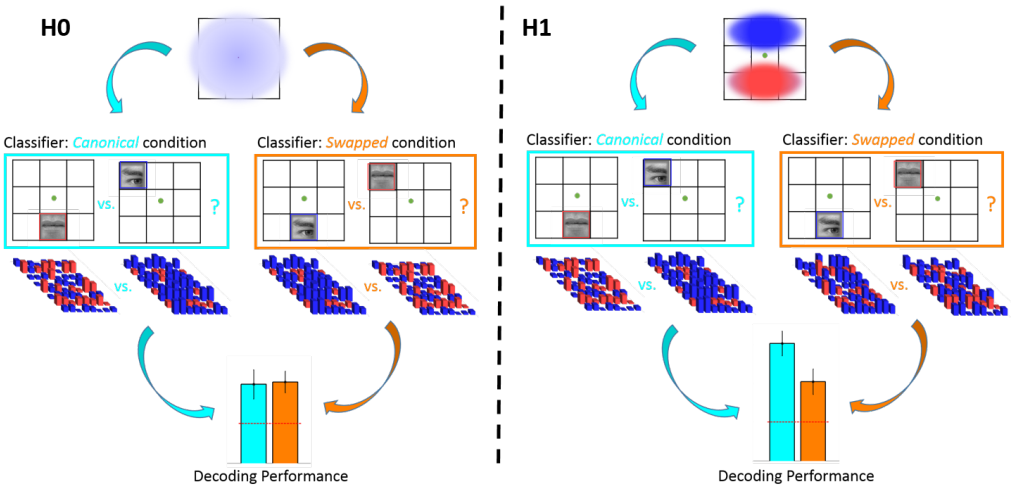

Figure 7-4 Hypotheses for multi-voxel pattern analyses. Left hand side: According to the null hypothesis voxel responses are location invariant. The population receptive fields (pRFs) for voxels with a preference for eye or mouth stimuli are identical (illustrated in purple and covering the entirety of the central visual field; top panel). Thus stimuli shown in *canonical* (cyan) vs. *swapped* (orange) locations will evoke identical responses (middle panel). The response patterns evoked by example eye and mouth stimuli are illustrated below for either condition. Response patterns evoked by the example eye stimulus are dominated by voxels preferring eye stimuli, shown in blue. The response pattern evoked by the example mouth stimulus are dominated by mouth preferring voxels, shown in red. The difference between patterns is similar in both conditions and consequentially decoding performance is similar between conditions (lower panel). **Right hand side**: According to the alternative hypothesis voxels with a preference for eye and mouth stimuli also have a spatial preference for the upper and lower visual field, respectively (top panel; pRFs shown in blue and red, respectively). Thus stimuli shown in *canonical* (cyan) vs. *swapped* (orange) locations will evoke different response patterns (middle panel). Specifically, response patterns evoked by eye stimuli will be dominated *more strongly* by voxels preferring eye stimuli (blue) in the *canonical* condition. Similarly, response patterns evoked by mouth stimuli will be dominated more strongly by mouth preferring voxels (red) in the *canonical* condition. This is because in the canonical condition the respective voxels will be stimulated in a way corresponding to their preferences in either dimension (face feature and retinotopic location). Furthermore, in the *swapped* condition voxels with a preference for the face feature not shown in a given trial will be more likely to respond nonetheless because the stimulus will be shown at a location they prefer. Consequentially the difference between patterns evoked by eye and mouth stimuli is more marked in the *canonical* condition and decoding performance will be systematically better for this condition, according to the alternative hypothesis (lower panel).

7.3.1.5.3 Mass-univariate analysis

To test for differences in general amplitude levels an additional mass-univariate analysis was performed. For this a general linear model was specified for each participant, including data from all ten runs and regressors for each trial type and run as well as motion regressors for each run (see above). Additionally the model contained regressors for the intercept of each run.

To describe and compare general amplitude levels in either condition, two contrast images were derived for each participant. One corresponded to the contrast of trials from the *canonical* condition vs. baseline trials (*eye* stimuli presented in the *upper visual field*; *mouth* stimuli presented in the *lower visual field*). The other image corresponded to the contrast between all trial types in the *swapped* condition vs. baseline trials (*eye* stimuli presented in the *lower visual field*; *mouth* stimuli presented in the *upper visual field*). Additional analyses indicated that response amplitudes did not differ significantly for stimuli shown in the contra- and ipsi-lateral hemifield. Thus results reported here used contrasts collapsing across stimuli in either hemi-field for each condition.

For both contrast maps the average values within each ROI were calculated for each participant. Then these averages were compared against zero using one-sample *t*-tests and against each other using paired sample *t*-tests.

7.3.2 Results

7.3.2.1 Results from multi-voxel pattern analysis

Stimulus identity (eye *vs.* mouth stimuli) could be decoded significantly better than chance level in either condition and from all four ROIs – except from right FFA in the *swapped* condition. **Table 7-2** gives the mean decoding accuracies for each condition and ROI, their standard error and results of one-sample *t*-tests against chance level (50%).

The rightmost column of **Table 7-2** gives results of the comparison of decoding accuracies between conditions. Condition did not have a significant influence on decoding accuracy in the left hemisphere, but was significantly better for the canonical vs. swapped conditions in right IOG (t_{20}=2.20, P<.05) and there was a similar trend in right FFA (t_{20}=1.92, P=.07; c.f. **Figure 7-5**).

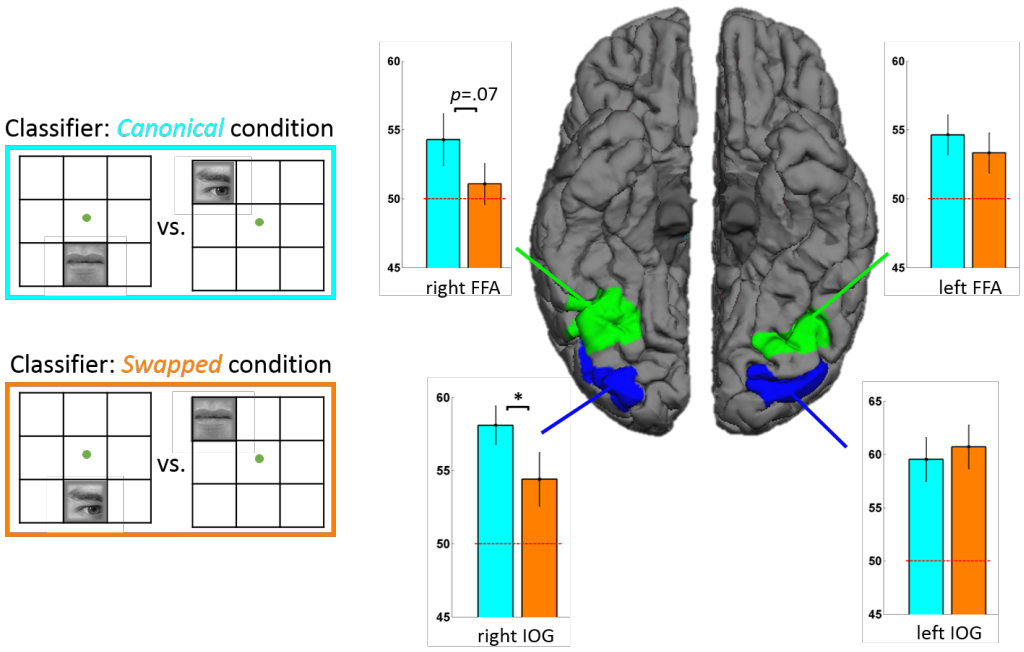

Figure 7-5 Classification performance by condition and region of interest. **Left hand side**: Illustration of conditions. In the *canonical* condition (cyan) mouth and eye stimuli were shown in typical lower and upper visual field locations (assuming fixation just above the nose). In the *swapped* condition (orange), these locations were swapped between eye and mouth stimuli (green dot indicates fixation; see Methods for details). **Right hand side**: Regions of interest (ROIs) and decoding results. The image depicts the inferior view of both hemispheres for an example participant (three dimensional rendering done using FreeSurfer). The inferior occipital gyrus (IOG) and fusiform face area (FFA) ROIs are shown in blue and green, respectively. Plots show decoding results for all four regions of interest and either condition. Values on the Y-axes indicate decoding performance in percent correct, bars and error bars indicate the mean and S.E.M. of decoding performance across participants. Plots correspond to regions as indicated by colored lines and labels below the x-axis. Condition is indicated by bar color (*canonical* condition in cyan, *swapped* condition in orange). The dashed horizontal red line highlights chance level. All decoding was significantly better than chance level, apart from right FFA and the *swapped* condition (orange bar in upper left plot). For data from the right IOG decoding was significantly better in the *canonical* vs. *swapped* condition (lower left plot) and a similar trend was seen for the right FFA (upper left plot).

Table 7-2 Classification performance by condition and region of interest. Cells contain the mean and standard error of the mean (S.E.M.) across participants for decoding performance of a linear support vector machine classifying whether patterns of BOLD activation in a region of interest were elicited by eye or mouth stimuli (see Methods for details). *t* and *P* values in the first two columns correspond to one-sample t-tests of decoding performance against chance level (50%). The right most column gives the mean and S.E.M. for a decoding advantage in the *canonical* vs. *swapped* condition (see Methods for details of conditions). *t* and *P* values in this column correspond to a paired-sample *t*-test comparing decoding performance between conditions. IOG: inferior occipital gyrus, FFA: fusiform face area. C.f. **Figure 7-5**.

	***Canonical* condition**	***Swapped* condition**	***Canonical* advantage**
Left IOG	59.52% (2.06%) t_{20}=4.62, *P*<.001	60.71% (2.04%) t_{20}=5.25, *P*<.0001	-1.19% (2.08%) t_{20}=-0.57, *P*=.57, *n.s.*
Left FFA	54.64% (1.44%) t_{20}=3.23, *P*<.01	53.33% (1.45%) t_{20}=2.30, *P*<.05	1.31% (2.00%) t_{20}=0.65, *P*=.52, *n.s.*
Right IOG	58.10% (1.31%) t_{20}=6.17, *P*<.00001	54.40% (1.83%) t_{20}=2.40, *P*<.05	3.69% (1.68%) t_{20}=2.20, *P*<.05
Right FFA	54.29% (1.88%) t_{20}=2.29, *P*<.05	51.07% (1.49%) t_{20}=0.72, *P*=.48, *n.s.*	3.21% (1.67%) t_{20}=1.92, *P*=.07, *n.s.*

7.3.2.2 Results from mass-univariate analysis

All four ROIs responded significantly to face feature stimuli in either condition (as compared to baseline trials). **Table 7-3** gives the mean response amplitudes for each condition and ROI (arbitrary units), the corresponding standard errors and results of one-sample *t*-tests against zero.

The rightmost column of **Table 7-3** gives results of the comparison of amplitudes between conditions. Condition did not have a significance influence on response amplitude in any of the ROIs (c.f. **Figure 7-6**).

Table 7-3 Response amplitude by condition and region of interest. Cells contain the mean and standard error of the mean (S.E.M.) across participants for BOLD amplitudes elicited by eye or mouth stimuli in a region of interest (compared to baseline trials; arbitrary units; see Methods for details). *t* and *P* values in the first two columns correspond to one-sample t-tests of response amplitude against zero. The right-most column gives the mean and S.E.M. for the difference in response amplitudes in the *canonical-swapped* conditions (see Methods for details of conditions). *t* and *P* values in this column correspond to a paired-sample *t*-test comparing response amplitudes between conditions. IOG: inferior occipital gyrus, FFA: fusiform face area. C.f. **Figure 7-6**.

	***Canonical* condition**	***Swapped* condition**	***Canonical* advantage**
Left IOG	111.20 (20.51) t_{20}=5.42, *P*<.0001	116.56 (22.21) t_{20}=5.25, *P*<.0001	-5.37 (5.70) t_{20}=-0.94, *P*=.36, *n.s.*
Left FFA	69.88 (14.51) t_{20}=4.82, *P*<.001	68.42 (16.08) t_{20}=4.25, *P*<.001	1.31% (2.00%) t_{20}=0.65, *P*=.52, *n.s.*
Right IOG	123.96% (17.40) t_{20}=7.13, *P*<.000001	128.78 (21.35%) t_{20}=6.03, *P*<.00001	1.46 (4.84) t_{20}=0.30, *P*=.77, *n.s.*
Right FFA	72.11 (13.07) t_{20}=5.52, *P*<.0001	72.46 (16.36) t_{20}=4.43, *P*<.001	-0.35 (5.43) t_{20}=-0.06, *P*=.95, *n.s.*

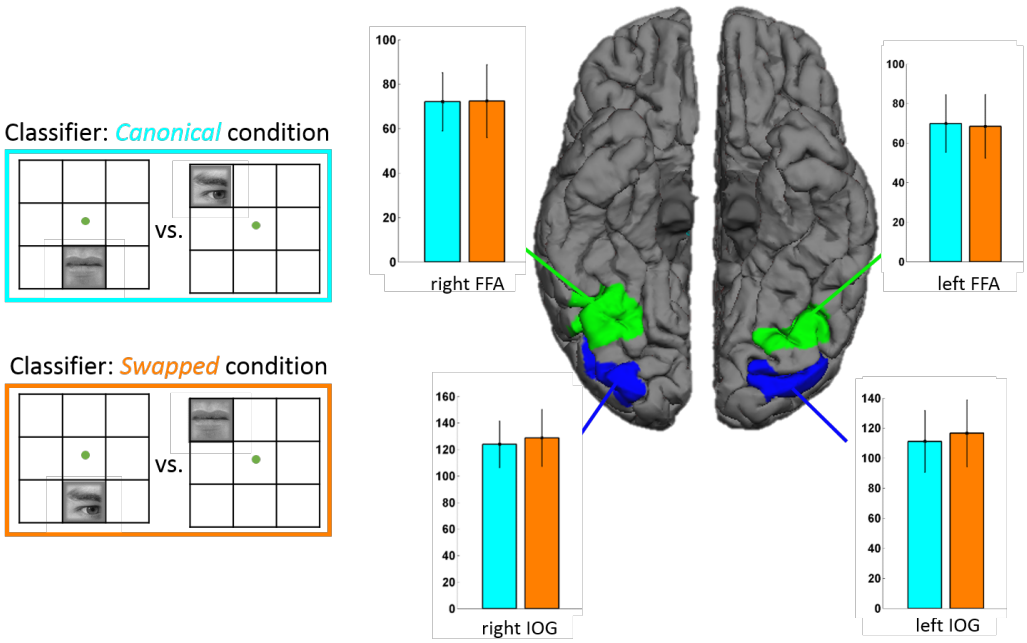

Figure 7-6 Response amplitude by condition and region of interest. Left hand side: Illustration of conditions. In the *canonical* condition (cyan) mouth and eye stimuli were shown in typical lower and upper visual field locations (assuming fixation just above the nose). In the *swapped* condition (orange), these locations were swapped between eye and mouth stimuli (green dot indicates fixation; see Methods for details). **Right hand side**: Regions of interest (ROIs) and corresponding response amplitudes. The image depicts the inferior view of both hemispheres for an example participant (three dimensional rendering done using FreeSurfer). The inferior occipital gyrus (IOG) and fusiform face area (FFA) ROIs are shown in blue and green, respectively. Plots show response amplitudes for all four regions of interest and either condition. Values on the Y-axes indicate response amplitudes elicited by face feature stimuli vs. baseline trials in arbitrary units; bars and error bars indicate the mean and S.E.M. across participants. Plots correspond to regions as indicated by colored lines and labels below the x-axis. Condition is indicated by bar color (*canonical* condition in cyan, *swapped* condition in orange). Response amplitude in all ROIs were significantly greater than zero in either condition but did not differ significantly between conditions (c.f. **Table 7-3**).

7.3.3 Discussion

IOG and FFA of both hemispheres responded to isolated face features and the mean amplitude of these responses did not differ for stimuli

presented at *canonical* and *swapped* locations. Moreover, the patterns of activation within each area allowed distinguishing between eye and mouth stimuli significantly better than chance level in either condition (except for right FFA, where stimulus decoding was only successful for the *canonical* condition). Crucially, the accuracy with which eye and mouth evoked patterns could be separated was significantly better for the *canonical* condition in right IOG and there was a similar trend in right FFA.

Significant responses to isolated face features were a pre-requisite for testing the main hypothesis. This finding is inline with previous reports of responses to partial face stimuli in fMRI (Chan et al., 2010; Kriegeskorte et al., 2013; Liu et al., 2010), EEG (Smith et al., 2004) and MEG (Smith et al., 2009). The fact that overall amplitude levels were equal between conditions is compatible with the null *and* the alternative hypotheses. Note that the hypotheses make different predictions only with regard to the *distribution* of response amplitudes within an area as a function of condition and stimulus type (c.f. **Figure 7-4**).

A second pre-requisite for the alternative hypothesis was that eye and mouth stimuli would elicit separable patterns of activation in face sensitive patches of cortex. This was the case for either region of interest, hemisphere and condition (apart from the *swapped* condition in right FFA). This suggests that voxels in IOG and FFA show some degree of tuning for face features with differential preferences for eye and mouth stimuli.

The alternative hypothesis predicted that face feature decoding would be better for *canonical* stimulus locations. Exactly this pattern of results has been found for right IOG (and possibly right FFA). The apparent

lateralization of this effect is in line with a general right hemisphere dominance for face processing (e.g. (Caspers et al., 2013; Dien, 2009; Willems, Peelen, & Hagoort, 2010)) and for the OFA specifically (Chan et al., 2010; Kovács et al., 2008; Pitcher et al., 2011, 2007). Also, the fact that the data provided stronger evidence for the predicted effect in IOG than FFA is consistent with previous results. (Right) OFA – in comparison to FFA - seems particularly sensitive to stimulus location (Kovács et al., 2008; Schwarzlose et al., 2008) and face parts (Liu et al., 2010; Nichols et al., 2010; Pitcher et al., 2007). Also note that (Issa & DiCarlo, 2012) propose OFA as a potential homologue for the posterior face sensitive patch of monkey IT for which they found systematic retinotopic and feature tuning (see above). In line with these previous results, some models of face processing propose the role of (right) OFA to be the early, structural encoding of faces (Haxby et al., 2000; Pitcher et al., 2011).

Taken together, previous results point to a role of stimulus location and feature processing in (right) OFA. The neuroimaging results presented here suggest an interaction between spatial and feature preferences in this area (and possibly in right FFA). An interaction of spatial and feature tuning properties is of potential relevance to understand the neural mechanisms of face perception, well established phenomena like the face inversion effect and, possibly, clinical conditions affecting face recognition. These and other points will be discussed in more detail below.

7.4 General chapter discussion

This chapter presented a behavioral and a neuroimaging experiment with converging results, pointing to an advantage for the processing of eye and mouth stimuli presented at canonical retinotopic locations. This, in combination with previous studies, is taken as evidence for a correlation between retinotopic and feature preferences in face sensitive neural populations (e.g. (Issa & DiCarlo, 2012)).

The effects of stimulus location on recognition performance were surprisingly strong. They were significant for both, eye and mouth stimuli albeit in (the predicted) opposite directions and consequentially the interaction effect was large. The interaction nature of the main hypothesis ensures that general stimuli and location preferences are controlled for. Stimuli and stimulus locations were identical between conditions – it was only their *combination* that varied between conditions. I therefore interpret the behavioral result as strong evidence for a recognition advantage for face features presented at canonical stimulus locations. However, the current experiment tested only three stimulus locations. Future studies should sample the visual field more densely. This way one could systematically map perceptual sensitivity to face features and compare such maps between features (and potentially to respective maps of neural sensitivity, see below). Also, individual sensitivity maps could be tested for a link to individual differences in gaze behavior towards faces (c.f. (Peterson & Eckstein, 2013a)).

The neuroimaging results significantly support the main hypothesis and converge with both, the findings from the behavioral experiment as well

as previous literature. However, the statistical effect size of the neuroimaging result was weaker than that of the behavioral data. Several factors could enhance the power of future studies.

Stimuli were presented only very briefly (200 ms) and followed by a mask to ensure comparability of the behavioral and neuroimaging experiments. However, longer stimulus durations might go along with a stronger fMRI signal and avoid possible nonlinear effects of neurovascular coupling for brief stimulus durations (the linear time-invariant assumption can be violated for very brief stimulus durations; c.f. Chapter 2.5.3.2 and e.g. (Pfeuffer et al., 2003)). Further, longer localizer runs could enhance the power for detecting face sensitive patches of cortex and thus potentially enable the use of functionally defined, more specific ROI masks for OFA. Moreover, in the current study I acquired data from a rather small selection of slices centered on the inferior temporal cortex to maximize the amount of data from this area. Future studies might explore effects of stimulus location for the representation of face features in other regions, like posterior STS (e.g. (Freiwald et al., 2009; Perrett et al., 1982)) or the right inferior frontal junction (IFJ). According to a recent report IFG shows a preference for eye stimuli (regardless of whether they are embedded in a face or not; (Chan & Downing, 2011)).

Future studies could also more directly test the main hypothesis of co-varying tuning properties for retinotopic locations and face features. A way of doing this would be to fit parameters of encoding models to the time-series of each voxel (rather than using decoding algorithms across voxel patterns; e.g. (Dumoulin & Wandell, 2008; Kay, Naselaris, Prenger, & Gallant,

2008); Chapter 6). For instance, one could compare the retinotopic tuning profiles of voxels for different face features used as mapping stimuli. This technique could potentially reveal tuning properties of neural populations on the sub-voxel level. Ideally, future studies will also probe the tuning properties of face sensitive neural populations electro-physiologically. Patients with temporal epilepsy might provide a rare window of opportunity for this (Parvizi et al., 2012).

Finally, what are the implications of tuning for canonical retinotopic locations of face features? First of all, this finding might be relevant for understanding the neural mechanisms of face processing. It provides evidence for a combination of feature and location tuning that appears to reflect input statistics. This observation is in line with models of object recognition that propose a combination of tuning properties to achieve representations of object identity, location and structure ((Edelman & Intrator, 2000; Ullman, 2007); c.f. (Kravitz et al., 2008)). A reflection of input statistics appears useful as it potentially avoids a waste of resources – predictable retinotopic locations render a complete sampling mosaic of face features across the visual field excessive. The incorporation of such retinotopic 'priors' might also be seen as support for the idea of a 'Bayesian brain', aiming to minimize surprise (Friston, 2009). Future studies could investigate in how far such priors can be altered by training ((Peterson & Eckstein, 2013b); c.f. (Schmalzl, Palermo, Green, Brunsdon, & Coltheart, 2008)).

Retinotopic priors for face features might also shed new light on the well-established recognition advantage for upright over inverted faces (an

advantage that is larger for faces than other object categories; (Schwaninger et al., 2003; Valentine, 1988; Yin, 1969). Previous studies suggest that face inversion predominantly affects the recognition of configural aspects of faces (Leder & Bruce, 2000; Schwaninger & Mast, 2005), specifically the accurate perception of vertical distances between features (Goffaux & Rossion, 2007). However, the current results suggest that atypical retinotopic feature locations might play a role as well (see (Civile, McLaren, & McLaren, 2013) and (James et al., 2013) for the effects of inverting individual features). The eye and mouth features of an inverted face can never be at canonical locations simultaneously.

In this context it is worth noting the results of a recent behavioral and fMRI study by (James et al., 2013). Participants in this study observed upright or inverted low-contrast images of faces and isolated face features or feature combinations. Performance in a 1-back task was significantly affected by inversion only when stimuli contained features from the upper *and* lower half of the face and inversion of an isolated eye or pair of eyes had no effect. (James et al., 2013) presented stimuli for 1s each and in blocks according to stimulus type (e.g. inverted images containing an isolated eye) and report no fixation instruction. Thus it appears likely that participants foveated isolated features in this study – and thus escaped violations of canonical retinotopic locations for isolated features but not for those images containing feature combinations spanning the upper and lower halves of the face. Regarding the fMRI results, this study found no significant modulation of BOLD responses by inversion of isolated face features in OFA and FFA,

paralleling the results presented here (but note a significant inversion effect was reported for whole faces in FFA).

Furthermore, eye-tracking studies suggest that participants (at least initially) tend to look at the center of inverted faces (in a similar way as for upright faces; (Hills, Sullivan, & Pake, 2012)). So effectively inverted faces will be akin to the *swapped* condition used here. Interestingly, face inversion effects can be diminished by guiding observers' gaze towards the eye region (Hills, Cooper, & Pake, 2013; Hills, Ross, & Lewis, 2011). More generally, retinotopic priors for face features might aid the recognition of face-like first-order relations in an image (two eyes above a mouth) and thus the initial recognition of a face as a face (Maurer et al., 2002). Future studies could probe a link between feature position and face inversion effects directly by correlating the size of either effect across observers.

A final, more speculative implication of retinotopic feature preferences is their potential relevance for clinical conditions affecting face perception. Individuals with autism spectrum disorders (ASD) show at least moderate impairments in face recognition ((Barton et al., 2004; Boucher, Lewis, & Collis, 1998; Dawson, Webb, & McPartland, 2005; Klin et al., 1999; Weigelt, Koldewyn, & Kanwisher, 2012; Wilson, Palermo, & Brock, 2012); but see (Jemel, Mottron, & Dawson, 2006)). Furthermore, they tend to avoid looking at faces (Osterling, Dawson, & Munson, 2002), in particular the eye region ((Jones, Carr, & Klin, 2008; Jones & Klin, 2013; Klin, Jones, Schultz, Volkmar, & Cohen, 2002; Pelphrey et al., 2002; Tanaka & Sung, 2013; Wilson et al., 2012); but see e.g. (McPartland, Webb, Keehn, & Dawson, 2011)). Future studies could investigate whether this difference in gaze behavior is linked

to altered retinotopic priors for face features in ASD. Similarly, individuals with congenital prosopagnosia show atypical gaze behavior towards faces (Schwarzer et al., 2007). Congenital prosopagnosia is a hereditary deficit in face recognition of yet unclear neurobiological origin (Behrmann, Avidan, Marotta, & Kimchi, 2005; Grueter et al., 2007; Grüter, Grüter, & Carbon, 2008). Specifically, congenitally prosopagnosic individuals spread fixations across the whole face; unlike controls they tend to fixate the chin and hairline (Schwarzer et al., 2007). Although results regarding acquired prosopagnosia are more mixed, a similar pattern has been reported for at least one case (Stephan & Caine, 2009). This suggests that congenital prosopagnosia might go along with altered retinotopic priors for face features. Future studies should investigate whether individuals with ASD and congenital prosopagnosia differ from controls in paradigms like the one used here (see (Bate, 2013) for a detailed discussion of disorders affecting face recognition).

Interestingly, some studies report diminished or absent face inversion effects for individuals with congenital prosopagnosia (e.g. (Behrmann et al., 2005)) and ASD ((McPartland, Dawson, Webb, Panagiotides, & Carver, 2004; Rose et al., 2007); but see (Weigelt et al., 2012)). This could be related to altered retinotopic priors for face features as well (see above).

8 General Discussion and Conclusions

8.1 Introduction and summary of findings

In this chapter I will discuss general weaknesses and strengths as well as implications of the experiments presented earlier. More detailed discussions of individual findings and of the relevant literature can be found in the respective chapters. The common aim of experiments presented in this thesis was to investigate contextual modulations of visual perception and of visual cortex activity. Each chapter found evidence for the modulatory capacity of some such contextual factor.

The first three experimental chapters highlighted the role of cross-modal, auditory modulation in vision. In Chapter 3 the perceived duration of brief visual stimuli was found to vary with the duration of co-occurring sounds – at least as long as the duration differences between visual and auditory stimuli were only subtle and sounds were presented sufficiently close in time to the visual stimulus (Chapter 3, experiments 1 & 2). Although somewhat ambiguous, the results of an additional experiment (Chapter 3, experiment 3) suggest that auditory stimulus duration might alter the actual duration of the visual percept and thereby modulate the perception of non-temporal visual stimulus qualities. Chapter 4 replicated the sound-induced flash illusion and found that individual differences in proneness to this illusion are predicted by grey matter volume in early visual cortex. Participants who had a smaller portion of overall grey matter dedicated to these areas perceived the illusion systematically more often than those with larger visual cortices. The final cross-modal Chapter 5 applied decoding

algorithms to patterns of activity evoked in retinotopically defined areas V1-3. The accuracy with which dynamic visual stimuli could be decoded from V2 and V3 significantly decreased if these stimuli were paired with miss-matching sounds. Taken together, the first three experimental chapters provide evidence that sounds can alter visual perception (Chapters 3 and 4) as well as neural activity in early visual cortex (Chapter 5) and that the propensity for such effects depends on visual cortex anatomy (Chapter 4).

The following experimental Chapter 6 found that attention modulates the spatial tuning of neural population responses in V1-3. Specifically, task associated load at fixation modulated the spatial tuning of visual cortex for the surround. Spatial sensitivity was shifted away from the immediate surround of task stimuli and the representation of task-irrelevant stimuli in this area became spatially coarser. Interestingly, population receptive fields of voxels in intraparietal sulcus (IPS) showed the opposite pattern and were drawn *towards* the center of fixation. The findings of this chapter provide evidence that attention can modulate a fundamental aspect of the functional organization of V1-3, namely the mapping of the visual field onto the cortical surface.

In the final experimental Chapter 7 I tested perceptual and neural (population) sensitivity for face features as a function of retinotopic stimulus location. Observers showed a significant recognition advantage for face features presented at canonical retinotopic locations. Furthermore, patterns of fMRI activation in right OFA carried significantly more stimulus related information for face features presented at canonical (vs. swapped) locations. A similar trend was observed for activation patterns in right FFA.

This finding points to a significant role of spatial heterogeneity and retinotopic priors for the perception of face features.

Altogether, the results of the experiments presented in this thesis demonstrate the importance of contextual modulation for neural visual processing and its outcome, visual perception. Co-occurring sounds, attentional load and canonical stimulus locations were all found to alter visual processing and perception in subtle but reliably detectable ways.

8.2 Comparison of findings and first conclusions

Comparing the different modulatory effects described in this thesis with each other a broad pattern emerges. Sounds affected vision in a manner that was spatially coarse but temporally specific while the reverse was true for the effects of attentional load and (possibly) canonical stimulus locations.

Sounds specifically affected perceived visual duration and only did so when the temporal onsets of auditory and visual stimuli were closely matched (Chapter 3). At the same time this effect was robust against a high degree of spatial misalignment of sound and light sources (visual stimuli were presented on the vertical meridian while speakers were positioned next to the screen with a horizontal displacement of more than 20 degrees visual angle). The sound-induced flash illusion hinges upon a quick succession of events and close temporal proximity of flashes and beeps (Shams, Kamitani, & Shimojo, 2002), but was found to be constant across upper and lower visual field locations in Chapter 4. Finally, mismatching

sounds had detrimental effects on stimulus decoding from V2 and V3. This effect was interpreted to reflect temporal aspects of attention in Chapter 5 (c.f. (Large & Jones, 1999); although this idea has not been tested explicitly here).

Such temporally specific, spatially broad effects of auditory modulation contrast with the effect attentional load had on visual cortex representations. Attentional load explicitly altered the *spatial* tuning properties of neural populations in visual cortex. Moreover, this effect was spatially *specific* (i.e. varying with eccentricity). At the same time the effects of attentional load appear temporally coarse – they were detectable using a task lasting several minutes at a time (Chapter 6) and do not hinge upon exact temporal co-incidence of task related and distracting stimuli (Carmel et al., 2011)[5]. Finally, the findings of Chapter 7 show that retinotopic stimulus location can modulate processing even in 'higher' visual areas in a non-trivial[6] way.

So the first observation from my experiments is that sound had spatially coarse but temporally specific effects on visual perception. As noted in the General Introduction, this observation matches resolution differences between these modalities (e.g. (Welch et al., 1986)). Auditory perception comes with higher temporal but lower spatial resolution than vision (e.g. (Grondin, 2010; Witten & Knudsen, 2005)). And while the multisensory experiments in this thesis focused on auditory modulation of

[5] But note that high perceptual load can reduce the flicker fusion threshold and thus reduce visual temporal resolution (Carmel et al., 2007)
[6] An example of 'trivial' modulation would be differences between visual processing for foveal and peripheral presentation of stimuli

visual processing, visual stimuli can vice versa dominate spatial judgments regarding audiovisual stimuli (e.g. (Pick et al., 1969)). Such a match between sensory precision for an encoded quality and multisensory dominance can serve perceptual judgments that are statistically optimal. That is, it can serve perceptual judgments that reflect a weighting of input streams that is proportional to their respective reliabilities (e.g. (Ernst & Banks, 2002)). There is some evidence that this is the case for the sound induced flash illusion as well. When the reliability of judgments for either modality is varied and measured in unimodal conditions (by altering the number of events), this modality-specific reliability predicts the degree of cross-modal bias in exactly the way a Bayes-optimal model would predict (Shams, Ma, et al., 2005). These and other observations fit with a general theoretical framework of cue-integration as a Bayes-optimal process. According to this view different sources of evidence are integrated in a way that optimally serves the goal to reliably infer the causes of a sensory event (e.g. (Körding et al., 2007; Yuille & Buelthoff, 1996)). This hypothesis also has implications for cortical function and architecture. Just as on the behavioral level, the degree to which cortical representations in one modality are integrated with corresponding representations from another modality should depend on their relative levels of precision or reliability. An auxiliary hypothesis stemming from this would be that the number of modulatory feedback and cross-modal synapses a sensory neuron receives should scale with its encoding precision (see below).

Interestingly, such integration appears to result in irreversible fusion of the underlying sensory streams (at least in some cases). So an

improvement in sensitivity for *distal* causes (e.g. 'How long did the audiovisual event last?' or 'How many rapid audiovisual events took place?') seems to come at the cost of low discriminability of *proximal* causes (e.g. 'Did the flash or the beep last longer?', 'Was the number of auditory and visual events the same?'). This can be understood as another aspect of the same general estimation process. In an ideal observer the degree to which cues are integrated should be highest when the estimated probability of joint causes is highest and vice versa. So, for instance, the estimated duration of a flash should be biased by a co-occurring sound (because of the higher temporal precision for auditory perception) – but only to the degree to which sound and flash appear to be part of the same external even. This is the case for Bayes-optimal models and seems to apply to human observers as well (Körding et al., 2007; Sato, Toyoizumi, & Aihara, 2007). And it matches with observations from the experiments presented here. For instance, sound duration was only found to bias perceived visual duration when visual and auditory stimuli had sufficiently similar onset times and durations (Chapter 3). Finally, this kind of behavior will be adaptive as long as such integration follows heuristics that match the statistics of the environment. An example of such a heuristic could be 'Auditory and visual events originating close in time and space usually have a common source'.

A Bayesian framework of cue integration might provide a useful perspective on contextual modulations of vision in general. The two non-auditory forms of modulation described in Chapters 6 and 7 could be understood as the integration of spatial priors in such a model. For instance, attentional load reduces spatial precision for the representation of task-

irrelevant parts of the visual field (Chapter 6) and presumably increases spatial precision for task related representations at the same time (just as spatial attention does; e.g. (Yeshurun & Carrasco, 1998, 1999)). Because the nature of the load task entails clear information about where in the visual fields relevant events are to be expected, such re-allocation of resources could be framed as the effects of changing spatial priors. Similarly, greater sensitivity for face features presented at canonical retinotopic locations could be understood as spatial priors for these features – either hard-wired or acquired through ontological learning.[7] Future experiments could test whether the integration of such spatial priors follows an ideal observer model. For instance, the location of task-related stimuli in a load paradigm could vary pseudo-randomly following a two-dimensional Gaussian distribution centered on fixation. Manipulating the width of this distribution would correspond to a manipulation of the precision of the corresponding prior in an ideal observer model. Would the load effect observed in Chapter 6 decline with increasing width of this distribution in a way predicted by such a model?

Taken together, the results presented in this thesis fit with the idea that contextual modulations of vision reflect the efficient use of limited resources to infer relevant[8] causes of sensory events. The general nature of this first conclusion provides a stimulating guideline for auxiliary

[7] Future experiments could test this idea by morphing different features (like eye and mouth) into each other and test their discrimination thresholds across the visual field. If there is indeed a spatially varying prior for either type of feature the corresponding thresholds should vary accordingly.

[8] Relevance here can depend for instance on the task at hand.

hypotheses and further research. But as with many such conclusions it appears *too* broad and vague to be strictly falsifiable itself.

8.3 Weaknesses and strengths

The ultimate goal of cognitive neuroscience is to explain how cognitive phenomena such as perception arise from neural processes (Churchland & Sejnowski, 1988). A general strategy for doing so is to experimentally vary cognitive states and probe resulting changes in brain states or vice versa. In the context of this thesis, this translates to relating contextual modulations of visual perception to co-occurring modulations of neural activity in visual cortex. I would like to argue that those findings that establish such a link are the most convincing and interesting ones in this thesis. At the same time, the lack or indirect nature of such a link points to the most promising future experiments to follow up some of the experiments.

For example, the observation that sound duration can modulate perceived visual duration directly inspires hypotheses regarding the underlying neural mechanisms. Slightly prolonged sounds might induce a sustained activation of neurons in visual areas. This would also explain an increased signal to noise ratio for other, non-temporal, visual stimulus properties. The results of experiment 3 in Chapter 3 suggest that detectability of a flash in a two-interval forced choice paradigm is enhanced when co-occurring sounds are prolonged. Importantly, the sounds and their duration are not informative about which of the intervals contain the flash. Thus, prolonging sound duration has an effect on visual processing that is

not limited to duration-specific aspects, just as sustained activation of relevant visual neurons might. In a conceptually similar experiment (Berger et al., 2003) found that the sound-induced, illusory re-appearance of a tilted Gabor-patch enhances sensitivity for the orientation of the stimulus. Again, these findings cannot be explained by an interpretation that holds sounds would modulate temporal aspects of visual perception in a way that is specific to duration perception. These results can and should inform hypotheses about the neural processes underlying such auditory modulation effects. They render an effect of sounds on neural activity in visual areas more likely than an (exclusive) modulation of neural activity in (hypothetical) supra-modal areas specialized for duration perception. For example, prolonged sound duration might increase the amplitude levels of activation in visual areas or prolong recurrent activation of these neural populations.

This idea could be tested in MEG or EEG studies investigating early and late components of visually evoked potentials over occipital cortex. Do they vary as a function of co-occurring sound durations? Additionally, such studies could aim at decoding low level visual attributes of briefly flashing stimuli (like the orientation of Gabor patches and possibly from data restricted to occipital sources). A sliding-window approach would allow an investigator to describe the time-course of information representation in the electrophysiological data (e.g. (Ramkumar, Jas, Pannasch, Hari, & Parkkonen, 2013)) and to test whether and how it depends on the duration of co-occurring sounds. Furthermore, Dynamic Causal Modelling (DCM; (Daunizeau, David, & Stephan, 2011; Friston, Harrison, & Penny, 2003))

would allow comparison of different models of cortical interactions underlying such modulatory effects. While it is difficult to distinguish between direct and indirect effects of connectivity, DCM would allow to evaluate and compare models posing direct cross-talk between auditory and visual cortices alone versus feedback from superior temporal sulcus (or both).

Finally, the behavioral importance of these (hypothesized) neural effects could be tested directly using transcranial magnetic stimulation (TMS). For example, one might hypothesize that effects of prolonged sounds on visual duration perception depend on recurrent activation of visual cortices via feedback from posterior STS. This would allow for some concrete predictions regarding the effects of single pulse TMS. Generally, this hypothesis would predict that TMS over occipital cortex and posterior STS should both be able to disrupt the modulatory effect of sound. Crucially, it would predict different critical time-windows for the effects of TMS over either area. In an experiment like experiment 3 in Chapter 3, TMS over occipital cortex coinciding with a first feed-forward sweep of activation should disrupt detection performance regardless of co-occurring sound duration (or indeed whether or not a sound is present). In contrast, TMS over posterior STS should specifically disrupt the sound induced modulation of detection performance. Finally, the effect of TMS over occipital cortex would be predicted to wane after the first feed forward sweep but re-appear in a second critical time-window corresponding to re-current activations and extending beyond the end of the corresponding window for TMS over posterior STS (c.f. (Pascual-Leone & Walsh, 2001)).

The finding that individual proneness to the sound induced flash illusion is linked with visual cortex size is one step closer to providing a causal link between behavior and neural processes. At the very least correlations between individual differences in neuroanatomy and behavior enable informed speculation about underlying neural mechanisms (Kanai & Rees, 2011). Tracer studies in macaque provide evidence for direct feedback connections from auditory and multisensory areas to V1 and V2 (Clavagnier et al., 2004; Falchier et al., 2002; Rockland & Ojima, 2003). In humans, probabilistic tractography shows white matter tracts between auditory and visual cortex (Beer et al., 2011). Functional studies recording from neurons in early visual cortex of cats and rodents suggest that such feedback connections have the capacity for subthreshold modulation of 'unimodal' visual neurons (that cannot be driven by auditory input alone; (Allman & Meredith, 2007; Allman et al., 2008, 2009; Iurilli et al., 2012)). In light of these results it is tempting to speculate that relatively smaller visual cortices are associated with stronger auditory modulation of visual perception for reasons associated with the density of these feedback connections. If the number of relevant synapses predominantly scales with the number of neurons in areas from which they arise then smaller visual cortices should receive more cross-modal and feedback projections per neuron (**Figure 8-1**). This speculation could be tested. For instance, future tracer studies in macaque could target individuals with particularly large and small visual cortices as determined via retinotopic mapping. The hypothesis would predict that despite the difference in visual cortex size similar numbers of cross-modal connections ending in early visual cortex should be found. In

humans, a combination of retinotopic mapping and probabilistic tractography (e.g. (Beer et al., 2011)) could give a quantitative handle on individual visual cortex size, (rough) cross-modal and feedback connection density and the hypothesized relationship between the two.

A related idea is that the density of cross-modal and feedback connections might be inversely related to cortical magnification. Just as feedback density might be 'diluted' for overall bigger visual cortices (**Figure 8-1**), 'expanding' the patch of cortex representing a given part of the visual field could have a similar effect. This idea could be tested in the same kinds of experiments outlined above. Additionally, this hypothesis would predict a systematic relationship between cortical magnification for e.g. a given eccentricity band and illusion proneness for the respective visual field location. This could be tested by quantifying illusion propensity for different eccentricities and acquiring retinotopic maps of the same individuals.

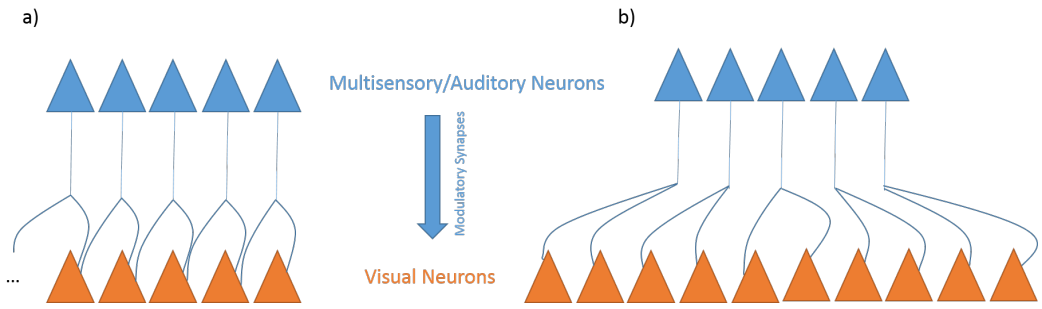

Figure 8-1 Schematic diagram of hypothesized link between visual cortex size and number of feedback connections. The number of feedback synapse from multisensory / auditory neurons (shown in blue) is assumed to (predominantly) scale with the number of neurons in areas from which they arise (two descending synapses per neuron in the schema shown). In **a)** each visual neuron receives two such synapses. In **b)** the number of visual neurons is doubled and thus the number of feedback connections per visual neuron is reduced to 1.

Stimulus decoding from V2 and V3 indicated that incongruent sounds introduce noise to early visual cortex representations of dynamic stimuli. However, the size of this effect was moderate and it is unclear whether it bears perceptual relevance. Even though mismatching sounds introduced measurable noise in early visual cortex representations, they presumably had no readily noticeable influence on visual perception of the presented video clips. Sounds can have drastic effects on visual perception (e.g. Chapter 4), but these effects appear to hinge upon a low signal to noise ratio for the respective visual representations. That is, only the perception of relatively weak visual stimuli appears to be modulated by sounds (e.g. (Bolognini et al., 2005); but see (Holmes, 2009)). The stimuli used in Chapter 5 in contrast were well visible and not designed to probe perceptual thresholds in any way. A potential way of increasing the perceptual relevance of sound congruency in this paradigm would be to present degraded stimuli. The design could be modified to systematically vary noise levels for visual stimuli and include perceptual judgments. This would allow establishing a more direct link between perceptual and neural measures of representational noise (and has been done in a similar way by (von Saldern & Noppeney, 2013)).

Similarly, the finding that attentional load can induce changes in the visual field maps of V1-3 needs to be linked to perceptual consequences. Such a link could shed light on the observation that attentional load had opposite effects on population receptive fields (pRFs) of early areas (V1-3) and of intraparietal sulcus (IPS). While pRFs representing the surround in

V1-3 were repelled from the target location at fixation, pRFs of IPS were drawn closer to the target. This pattern points to an interesting question: Which 'lines' (if any) are 'labeled' for perception (c.f. (Anton-Erxleben & Carrasco, 2013))? Put another way, does retinotopy merely reflect the current inputs of a neural population or does it reflect its functional role for localization (and possibly the representation of structure)? The answer to this question might differ between cortical areas and for represented stimulus qualities. Carefully planned psychophysical experiments could shed light on this. For instance, if the baseline retinotopy of a given area, voxel or neuron functions as ground truth for perceptual localization, then a shift of its receptive field should go along with perceptual mis-localization (c.f. (Anton-Erxleben & Carrasco, 2013)). Further, if perceptual load at fixation would indeed go along with mis-localization of peripheral stimuli, then the direction of this error could decide between V1-3 versus IPS as the more likely candidate for carrying such 'labels'. Finally, the observed changes in spatial tuning systematically varied across the visual field. Future experiments could test whether distractor suppression (and perceptual blurring? c.f. (Yeshurun & Carrasco, 1999)) reflect this heterogeneity across the visual field.

The final experimental chapter progressed from a biologically inspired hypothesis via a behavioral prediction and experiment to a neuroimaging study. The results of these experiments converged to suggest a role of canonical retinotopic locations for the perception of face features. Recognition performance for such features depended on location and so did their representations in face sensitive cortical areas of the right hemisphere.

Furthermore, these findings mirror recent results showing that eye-preferring cells of the posterior face sensitive patch in macaque have receptive fields at retinotopic locations corresponding to canonical eye positions (Issa & DiCarlo, 2012). Future experiments could aim to more directly test feature and spatial preferences of neurons in human face sensitive areas, e.g. by using electrocorticography in temporal lobe epilepsy patients (c.f. (Parvizi et al., 2012)). Another approach would be to use population receptive field mapping (Dumoulin & Wandell, 2008) with different face features as mapping stimuli. The spatial distribution of responses of a given cortical area could be projected back into the visual field and the resulting maps of neural sensitivity could be compared across face features and to corresponding maps of *perceptual* sensitivity for the same features. This way the link between BOLD responses in a face sensitive cortical area and perception would be tested more closely. Additionally, the same approach might be able to link potential individual differences in perceptual heterogeneity and gaze behavior (c.f. (Peterson & Eckstein, 2013a)) to corresponding differences in functional neuroanatomy.

A final, general caveat is that the size of neural and perceptual effects found in the experimental chapters was generally rather small. Even statistically very strong effects like the recognition advantage for face features at canonical locations only accounted for a performance difference of a few percent. However, I would argue that this lies in the very nature of the topic. Contextual modulations of visual perception are rather subtle – observers do not turn blind when presented with mismatching sounds, for

instance. Still these effects can reveal attributes of the visual system that are of potential relevance for understanding its functions.

Nevertheless, small effect sizes underscore the necessity to relate neural and perceptual findings to each other. A generally promising approach seems to embrace individual differences between observers as a source of valuable data rather than 'averaging them out' (Kanai & Rees, 2011). Similarly, perceptual heterogeneity across the visual field (e.g. (Afraz et al., 2010)) provides the opportunity to link behavioral measures to heterogeneity in corresponding neural representations. From a general cognitive neuroscience perspective, visual cortex appears an ideal target system for such experiments. It is organized in visual field maps on a broad spatial scale that allows to probe their layout with fMRI (Wandell et al., 2007). The high signal to noise ratio of fMRI signals in these areas enables reliable estimates of tuning parameters even on the single voxel level (Dumoulin & Wandell, 2008). Finally, there are remarkable individual differences with regard to the functional layout of visual cortex (Dougherty et al., 2003) and they can potentially be linked to individual differences in perception (Schwarzkopf et al., 2011).

8.4 Final Conclusions

Visual perception and neuronal processing are not a pure function of presented stimuli. The findings of this thesis add some examples of how contextual factors can modulate vision. The fact that visual processing is modulated by sounds, attention and canonical stimulus location probably

hints to the adaptive pressures shaping vision. Rather than serving a one-to-one mapping of the environment, vision primarily appears a tool to infer the causes of sensory events (e.g. (Friston, 2010; Körding et al., 2007)). Under this premise the incorporation of contextual factors in the process appears useful rather than spurious.

A related and final general conclusion is that the visual system resembles London rather than Chicago (**Figure 8-2**). A historically grown, layered and sometimes confusing structure, smooth and flexible boundaries, shortcuts and a pragmatic attitude towards non-perfect material seem characteristic of visual cortex and its function more than a schematic layout with straight connections and perfectly strict segregation of components. Even though areas of visual cortex show clear functional specialization and appear organized in a recognizable hierarchy (e.g. (Riesenhuber & Poggio, 1999)), the communication between areas is far from unidirectional (e.g. (Van Essen, Anderson, & Felleman, 1992)) and their specialization might be less clear cut and exclusive than sometimes suggested (e.g. (DiCarlo & Cox, 2007; Kravitz et al., 2008)). In the experiments presented here I found representations in early visual cortex to be modulated by sounds as well as attentional load, and the nature of representations in inferior temporal cortex to depend on retinotopic stimulus location. This underscores how the visual systems functions as a reciprocally interconnected whole and the necessity to understand its function as such.

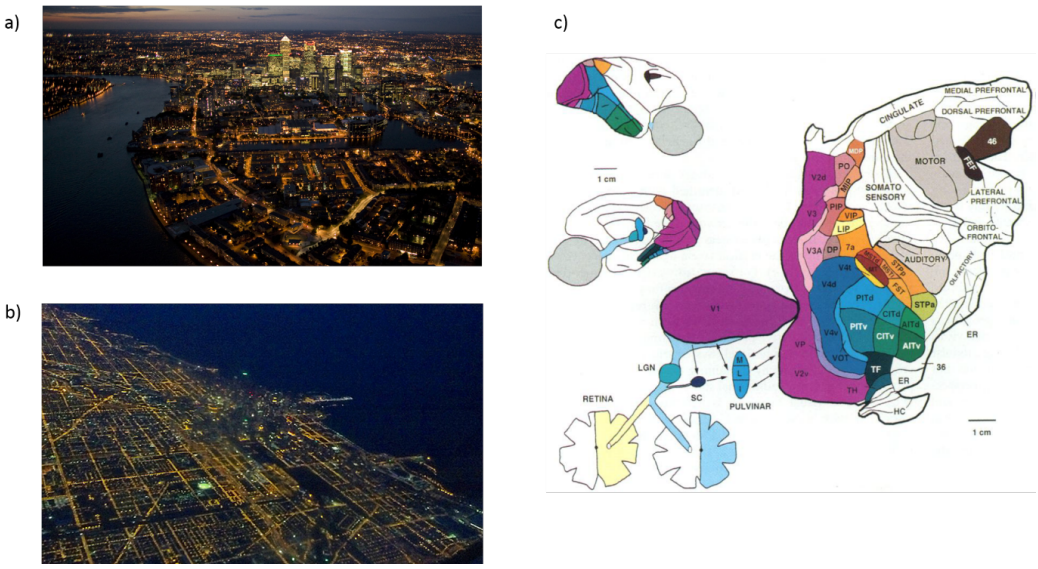

Figure 8-2 London, Chicago and the Visual System. **a)** London **b)** Chicago **c)** Overview of the macaque visual system as given by (Van Essen et al., 1992).

9 References

Abrams, J., Barbot, A., & Carrasco, M. (2010). Voluntary attention increases perceived spatial frequency. *Attention, Perception & Psychophysics*, *72*(6), 1510–21. doi:10.3758/APP.72.6.1510

Afraz, A., & Cavanagh, P. (2009). The gender-specific face aftereffect is based in retinotopic not spatiotopic coordinates across several natural image transformations. *Journal of Vision*, *9*(10), 10.1–17. doi:10.1167/9.10.10

Afraz, A., Pashkam, M. V., & Cavanagh, P. (2010). Spatial heterogeneity in the perception of face and form attributes. *Current Biology : CB*, *20*(23), 2112–6. doi:10.1016/j.cub.2010.11.017

Afraz, S.-R., & Cavanagh, P. (2008). Retinotopy of the face aftereffect. *Vision Research*, *48*(1), 42–54. doi:10.1016/j.visres.2007.10.028

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology : CB*, *14*(3), 257–62. doi:10.1016/j.cub.2004.01.029

Alink, A., Euler, F., Kriegeskorte, N., Singer, W., & Kohler, A. (2012). Auditory motion direction encoding in auditory cortex and high-level visual cortex. *Human Brain Mapping*, *33*(4), 969–78. doi:10.1002/hbm.21263

Allman, B. L., Bittencourt-Navarrete, R. E., Keniston, L. P., Medina, A. E., Wang, M. Y., & Meredith, M. A. (2008). Do cross-modal projections always result in multisensory integration? *Cerebral Cortex (New York, N.Y. : 1991)*, *18*(9), 2066–76. doi:10.1093/cercor/bhm230

Allman, B. L., Keniston, *Æ*. L. P., & Meredith, M. A. (2009). Not Just for Bimodal Neurons Anymore : The Contribution of Unimodal Neurons to Cortical Multisensory Processing. *Brain Topography*, 157–167. doi:10.1007/s10548-009-0088-3

Allman, B. L., & Meredith, M. A. (2007). Multisensory processing in "unimodal" neurons: cross-modal subthreshold auditory effects in cat extrastriate visual cortex. *Journal of Neurophysiology*, *98*(1), 545–9. doi:10.1152/jn.00173.2007

Anton-Erxleben, K., & Carrasco, M. (2013). Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. *Nature Reviews. Neuroscience*, *14*(3), 188–200. doi:10.1038/nrn3443

Anton-Erxleben, K., Henrich, C., & Treue, S. (2007). Attention changes perceived size of moving visual patterns. *Journal of Vision*, *7*(11), 5.1–9. doi:10.1167/7.11.5

Anton-Erxleben, K., Stephan, V. M., & Treue, S. (2009). Attention reshapes center-surround receptive field structure in macaque cortical area MT. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(10), 2466–78. doi:10.1093/cercor/bhp002

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*(1), 95–113. doi:10.1016/j.neuroimage.2007.07.007

Ashburner, J. (2009). Computational anatomy with the SPM software. *Magnetic Resonance Imaging*, *27*(8), 1163–1174. doi:10.1016/j.mri.2009.01.006

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry--the methods. *NeuroImage*, *11*(6 Pt 1), 805–21. doi:10.1006/nimg.2000.0582

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*(3), 839–51. doi:10.1016/j.neuroimage.2005.02.018

Ashby, F. G. (2011). *Statistical Analysis of FMRI Data* (p. 332). MIT Press.

Attwell, D., & Iadecola, C. (2002). The neural basis of functional brain imaging signals. *Trends in Neurosciences*, *25*(12), 621–5.

Bahrami, B., Carmel, D., Walsh, V., Rees, G., & Lavie, N. (2008). Unconscious orientation processing depends on perceptual load. *Journal of Vision*, *8*(3), 12.1–10. doi:10.1167/8.3.12

Bahrami, B., Lavie, N., & Rees, G. (2007). Attentional load modulates responses of human primary visual cortex to invisible stimuli. *Current Biology : CB*, *17*(6), 509–13. doi:10.1016/j.cub.2007.01.070

Barton, J. J. S., Cherkasova, M. V, Hefter, R., Cox, T. A., O'Connor, M., & Manoach, D. S. (2004). Are patients with social developmental disorders prosopagnosic? Perceptual heterogeneity in the Asperger and socio-emotional processing disorders. *Brain : A Journal of Neurology*, *127*(Pt 8), 1706–16. doi:10.1093/brain/awh194

Bate, S. (2013). *Face Recognition and its Disorders* (p. 241). Palgrave Macmillan.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*(7), 1391–7.

Beauchamp, M. S. (2005). Statistical criteria in FMRI studies of multisensory integration. *Neuroinformatics*, *3*(2), 93–113. doi:10.1385/NI:3:2:093

Beauchamp, M. S., Pasalar, S., & Ro, T. (2010). Neural substrates of reliability-weighted visual-tactile multisensory integration. *Frontiers in Systems Neuroscience*, *4*, 25. doi:10.3389/fnsys.2010.00025

Beer, A. L., Plank, T., & Greenlee, M. W. (2011). Diffusion tensor imaging shows white matter tracts between human auditory and visual cortex. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *213*(2-3), 299–308. doi:10.1007/s00221-011-2715-y

Behrmann, M., Avidan, G., Marotta, J. J., & Kimchi, R. (2005). Detailed exploration of face-related processing in congenital prosopagnosia: 1. Behavioral findings. *Journal of Cognitive Neuroscience*, *17*(7), 1130–49. doi:10.1162/0898929054475154

Ben Hamed, S., Duhamel, J.-R., Bremmer, F., & Graf, W. (2002). Visual receptive field modulation in the lateral intraparietal area during attentive fixation and free gaze. *Cerebral Cortex (New York, N.Y. : 1991)*, *12*(3), 234–45.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological Studies of Face Perception in Humans. *Journal of Cognitive Neuroscience*, *8*(6), 551–565. doi:10.1162/jocn.1996.8.6.551

Berger, T. D., Martelli, M., & Pelli, D. G. (2003). Flicker flutter: is an illusory event as good as the real thing? *Journal of Vision*, *3*(6), 406–12. doi:10:1167/3.6.1

Biederman, I., & Cooper, E. E. (1991). Evidence for complete translational and reflectional invariance in visual object priming. *Perception*, *20*(5), 585–93.

Biederman, I., Cooper, E. E., Kourtzi, Z., Sinha, P., & Wagemans, J. (2009). Biederman and Cooper's 1991 paper. *Perception*, *38*(6), 809–825.

Bisley, J. W. (2011). The neural basis of visual attention. *The Journal of Physiology*, *589*(Pt 1), 49–57. doi:10.1113/jphysiol.2010.192666

Blais, C., Jack, R. E., Scheepers, C., Fiset, D., & Caldara, R. (2008). Culture shapes how we look at faces. *PloS One*, *3*(8), e3022. doi:10.1371/journal.pone.0003022

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH '99* (pp. 187–194). New York, New York, USA: ACM Press. doi:10.1145/311535.311556

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially

converging audiovisual inputs. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *160*(3), 273–82.

Boucher, J., Lewis, V., & Collis, G. (1998). Familiar face and voice matching and recognition in children with autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *39*(2), 171–81.

Bouvier, S. E., & Engel, S. A. (2006). Behavioral deficits and cortical damage loci in cerebral achromatopsia. *Cerebral Cortex (New York, N.Y. : 1991)*, *16*(2), 183–91. doi:10.1093/cercor/bhi096

Bouvier, S. E., & Engel, S. A. (2011). Delayed effects of attention in visual cortex as measured with fMRI. *NeuroImage*, *57*(3), 1177–83. doi:10.1016/j.neuroimage.2011.04.012

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *16*(13), 4207–21.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–6.

Brefczynski, J. A., & DeYoe, E. A. (1999). A physiological correlate of the "spotlight" of visual attention. *Nature Neuroscience*, *2*(4), 370–4. doi:10.1038/7280

Bressler, D. W., & Silver, M. A. (2010). Spatial attention improves reliability of fMRI retinotopic mapping signals in occipital and parietal cortex. *NeuroImage*, *53*(2), 526–33. doi:10.1016/j.neuroimage.2010.06.063

Broadbent, D. E. (1952). Listening to one of two synchronous messages. *Journal of Experimental Psychology*, *44*(1), 51–55. doi:10.1037/h0056491

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review*, *64*(3), 205–15.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*(3).

Bruno, A., Ayhan, I., & Johnston, A. (2010). Retinotopic adaptation-based visual duration compression. *Journal of Vision*, *10*(10), 30. doi:10.1167/10.10.30

Budinger, E., Heil, P., Hess, A., & Scheich, H. (2006). Multisensory processing via early cortical stages: Connections of the primary auditory cortical

field with other sensory systems. *Neuroscience*, *143*(4), 1065–83. doi:10.1016/j.neuroscience.2006.08.035

Buhusi, C. V, & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews. Neuroscience*, *6*(10), 755–65. doi:10.1038/nrn1764

Burr, D., Banks, M. S., & Morrone, M. C. (2009). Auditory dominance over vision in the perception of interval duration. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *198*(1), 49–57. doi:10.1007/s00221-009-1933-z

Burton, A. M., Bruce, V., & Johnston, R. A. (1990). Understanding face recognition with an interactive activation model. *British Journal of Psychology*, *81*(3), 361–380. doi:10.1111/j.2044-8295.1990.tb02367.x

Buxton, R. B., Uludağ, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *NeuroImage*, *23 Suppl 1*, S220–33. doi:10.1016/j.neuroimage.2004.07.013

Cappe, C., Thelen, A., Romei, V., Thut, G., & Murray, M. (2012). Looming signals reveal synergistic principles of multisensory integration. *Journal of Neuroscience*, *32*(4), 1171–1182.

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2009). Selective integration of auditory-visual looming cues by humans. *Neuropsychologia*, *47*(4), 1045–52. doi:10.1016/j.neuropsychologia.2008.11.003

Cappe, C., Thut, G., Romei, V., & Murray, M. M. (2010). Auditory-visual multisensory interactions in humans: timing, topography, directionality, and sources. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(38), 12572–80. doi:10.1523/JNEUROSCI.1099-10.2010

Cardoso, M. M. B., Sirotin, Y. B., Lima, B., Glushenkova, E., & Das, A. (2012). The neuroimaging signal is a linear sum of neurally distinct stimulus- and task-related components. *Nature Neuroscience*, *15*(9), 1298–306. doi:10.1038/nn.3170

Carmel, D., Saker, P., Rees, G., & Lavie, N. (2007). Perceptual load modulates conscious flicker perception. *Journal of Vision*, *7*(14), 14.1–13. doi:10.1167/7.14.14

Carmel, D., Thorne, J. D., Rees, G., & Lavie, N. (2011). Perceptual load alters visual excitability. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(5), 1350–60. doi:10.1037/a0024320

Carrasco, M., Loula, F., & Ho, Y.-X. (2006). How attention enhances spatial resolution: evidence from selective adaptation to spatial frequency. *Perception & Psychophysics*, *68*(6), 1004–12.

Cartwright-Finch, U., & Lavie, N. (2007). The role of perceptual load in inattentional blindness. *Cognition*, *102*(3), 321–340.

Caspers, J., Zilles, K., Amunts, K., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2013). Functional characterization and differential coactivation patterns of two cytoarchitectonic visual areas on the human posterior fusiform gyrus. *Human Brain Mapping*. doi:10.1002/hbm.22364

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002). Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *Journal of Neurophysiology*, *88*(5), 2547–56. doi:10.1152/jn.00693.2001

Chan, A. W.-Y., & Downing, P. E. (2011). Faces and eyes in human lateral prefrontal cortex. *Frontiers in Human Neuroscience*, *5*, 51. doi:10.3389/fnhum.2011.00051

Chan, A. W.-Y., Kravitz, D. J., Truong, S., Arizpe, J., & Baker, C. I. (2010). Cortical representations of bodies and faces are strongest in commonly experienced configurations. *Nature Neuroscience*, *13*(4), 417–8. doi:10.1038/nn.2502

Chen, K.-M., & Yeh, S.-L. (2009). Asymmetric cross-modal effects in time perception. *Acta Psychologica*, *130*(3), 225–34. doi:10.1016/j.actpsy.2008.12.008

Chen, Y., Martinez-Conde, S., Macknik, S. L., Bereshpolova, Y., Swadlow, H. A., & Alonso, J.-M. (2008). Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nature Neuroscience*, *11*(8), 974–82. doi:10.1038/nn.2147

Chen, Y.-C., Huang, P.-C., Yeh, S.-L., & Spence, C. (2011). Synchronous sounds enhance visual sensitivity without reducing target uncertainty. *Seeing and Perceiving*, *24*(6), 623–38. doi:10.1163/187847611X603765

Churchland, P. S., & Sejnowski, T. J. (1988). Perspectives on cognitive neuroscience. *Science (New York, N.Y.)*, *242*(4879), 741–5.

Civile, C., McLaren, R. P., & McLaren, I. P. L. (2013). The face inversion effect-Parts and wholes: Individual features and their configuration. *Quarterly Journal of Experimental Psychology (2006)*. doi:10.1080/17470218.2013.828315

Clavagnier, S., Falchier, A., & Kennedy, H. (2004). Long-distance feedback projections to area V1: implications for multisensory integration,

spatial awareness, and visual consciousness. *Cognitive, Affective & Behavioral Neuroscience*, *4*(2), 117–26.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409–412. doi:10.3758/BF03203962

Colonius, H., & Diederich, A. (2011). Computing an optimal time window of audiovisual integration in focused attention tasks: illustrated by studies on effect of age and prior knowledge. *Experimental Brain Research*, *212*(3), 327–37. doi:10.1007/s00221-011-2732-x

Connor, C. E., Gallant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *Journal of Neurophysiology*, *75*(3), 1306–8.

Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(9), 3201–14.

Coull, J. T., Cheng, R.-K., & Meck, W. H. (2011). Neuroanatomical and neurochemical substrates of timing. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *36*(1), 3–25. doi:10.1038/npp.2010.113

Dahl, C. D., Logothetis, N. K., & Kayser, C. (2010). Modulation of visual responses in the superior temporal sulcus by audio-visual congruency. *Frontiers in Integrative Neuroscience*, *4*, 10. doi:10.3389/fnint.2010.00010

Daunizeau, J., David, O., & Stephan, K. E. (2011). Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *NeuroImage*, *58*(2), 312–22. doi:10.1016/j.neuroimage.2009.11.062

Dawson, G., Webb, S. J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: insights from behavioral and electrophysiological studies. *Developmental Neuropsychology*, *27*(3), 403–24. doi:10.1207/s15326942dn2703_6

De Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science (New York, N.Y.)*, *291*(5509), 1803–6. doi:10.1126/science.1056496

De Haas, B., Kanai, R., Jalkanen, L., & Rees, G. (2012). Grey matter volume in early human visual cortex predicts proneness to the sound-induced flash illusion. *Proceedings. Biological Sciences / The Royal Society*, *279*(1749), 4955–61. doi:10.1098/rspb.2012.2132

Deichmann, R., Schwarzbauer, C., & Turner, R. (2004). Optimisation of the 3D MDEFT sequence for anatomical brain imaging: technical

implications at 1.5 and 3 T. *NeuroImage*, *21*(2), 757–67. doi:10.1016/j.neuroimage.2003.09.062

Desimone, R., Albright, T., Gross, C., & Bruce, C. (1984). Stimulus-selective properties of inferior temporal neurons in the macaque. *J. Neurosci.*, *4*(8), 2051–2062.

Desseilles, M., Balteau, E., Sterpenich, V., Dang-Vu, T. T., Darsaud, A., Vandewalle, G., … Schwartz, S. (2009). Abnormal neural filtering of irrelevant visual information in depression. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(5), 1395–403. doi:10.1523/JNEUROSCI.3341-08.2009

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, *53*(1), 1–15. doi:10.1016/j.neuroimage.2010.06.010

Deutsch, J. A., & Deutsch, D. (1963). Some theoretical considerations. *Psychological Review*, *70*, 80–90.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., … Neitz, J. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(6), 2382–6.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences*, *11*(8), 333–41. doi:10.1016/j.tics.2007.06.010

DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology*, *89*(6), 3264–78. doi:10.1152/jn.00358.2002

Dickinson, J. E., Mighall, H. K., Almeida, R. A., Bell, J., & Badcock, D. R. (2012). Rapidly acquired shape and face aftereffects are retinotopic and local in origin. *Vision Research*, *65*, 1–11. doi:10.1016/j.visres.2012.05.012

Dien, J. (2009). A tale of two recognition systems: implications of the fusiform face area and the visual word form area for lateralized object recognition models. *Neuropsychologia*, *47*(1), 1–16. doi:10.1016/j.neuropsychologia.2008.08.024

Donovan, C.-L., Lindsay, D. S., & Kingstone, A. (2004). Flexible and abstract resolutions to crossmodal conflicts. *Brain and Cognition*, *56*(1), 1–4.

Dougherty, R. F., Koch, V. M., Brewer, A. a, Fischer, B., Modersitzki, J., & Wandell, B. a. (2003). Visual field representations and locations of

visual areas V1/2/3 in human visual cortex. *Journal of Vision*, *3*(10), 586–98. doi:10:1167/3.10.1

Driver, J., & Noesselt, T. (2008). Multisensory interplay reveals crossmodal influences on "sensory-specific" brain regions, neural responses, and judgments. *Neuron*, *57*(1), 11–23. doi:10.1016/j.neuron.2007.12.013

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–60. doi:10.1016/j.neuroimage.2007.09.034

Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron*, *38*(4), 659–71.

Eagleman, D. M. (2008). Human time perception and its illusions. *Current Opinion in Neurobiology*, *18*(2), 131–136.

Edelman, S., & Intrator, N. (2000). (Coarse coding of shape fragments) + (retinotopy) approximately = representation of structure. *Spatial Vision*, *13*(2-3), 255–64.

Eickhoff, S. B., Heim, S., Zilles, K., & Amunts, K. (2006). Testing anatomically specified hypotheses in functional imaging using cytoarchitectonic maps. *NeuroImage*, *32*(2), 570–82. doi:10.1016/j.neuroimage.2006.04.204

Eimer, M. (2011). The face-sensitive N170 component of the event-related brain potential. In A. Calder, G. Rhodes, M. Johnson, & J. Haxby (Eds.), *Oxford handbook of face perception*. Oxford: Oxford University Press.

Ekstrom, A. (2010). How and when the fMRI BOLD signal relates to underlying neural activity: the danger in dissociation. *Brain Research Reviews*, *62*(2), 233–44. doi:10.1016/j.brainresrev.2009.12.004

Ellis, R., Allport, D. A., Humphreys, G. W., & Collis, J. (1989). Varieties of object constancy. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *41*(4), 775–96.

Engel, S. A., Rumelhart, D. E., Wandell, B. A., Lee, A. T., Glover, G. H., Chichilnisky, E. J., & Shadlen, M. N. (1994). fMRI of human visual cortex. *Nature*, *369*(6481), 525. doi:10.1038/369525a0

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. doi:10.3758/BF03203267

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–33. doi:10.1038/415429a

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *22*(13), 5749–59. doi:20026562

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics*, *63*(4), 719–725. doi:10.3758/BF03194432

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*(1), 146–54. doi:10.1038/nn.2983

Fischer, J., & Whitney, D. (2009). Attention narrows position tuning of population responses in V1. *Current Biology : CB*, *19*(16), 1356–61. doi:10.1016/j.cub.2009.06.059

Fishman, M. C., & Michael, P. (1973). Integration of auditory information in the cat's visual cortex. *Vision Research*, *13*(8), 1415–9.

Fitousi, D., & Wenger, M. J. (2011). Processing capacity under perceptual and cognitive load: a closer look at load theory. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(3), 781–98. doi:10.1037/a0020675

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" Is Saying "What"? Brain-Based Decoding of Human Voice and Speech. *Science*, *322*(5903), 970–973.

Forster, S., & Lavie, N. (2008). Failures to ignore entirely irrelevant distractors: the role of load. *Journal of Experimental Psychology. Applied*, *14*(1), 73–83. doi:10.1037/1076-898X.14.1.73

Fortenbaugh, F. C., Prinzmetal, W., & Robertson, L. C. (2011). Rapid changes in visual-spatial attention distort object shape. *Psychonomic Bulletin & Review*, *18*(2), 287–94. doi:10.3758/s13423-011-0061-5

Frassinetti, F., Bolognini, N., & Làdavas, E. (2002). Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *147*(3), 332–43. doi:10.1007/s00221-002-1262-y

Freeman, E., & Driver, J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. *Current Biology*, *18*(16), 1262–1266.

Freiwald, W. A., & Tsao, D. Y. (2010). Functional compartmentalization and viewpoint generalization within the macaque face-processing system. *Science (New York, N.Y.)*, *330*(6005), 845–51. doi:10.1126/science.1194908

Freiwald, W. A., Tsao, D. Y., & Livingstone, M. S. (2009). A face feature space in the macaque temporal lobe. *Nature Neuroscience*, *12*(9), 1187–96. doi:10.1038/nn.2363

Friston, K. J. (2009). The free-energy principle: a rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. doi:10.1016/j.tics.2009.04.005

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews. Neuroscience*, *11*(2), 127–38. doi:10.1038/nrn2787

Friston, K. J. (2012). What does functional MRI measure? Two complementary perspectives. *Trends in Cognitive Sciences*, *16*(10), 491–2. doi:10.1016/j.tics.2012.08.005

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, *6*(3), 218–29. doi:10.1006/nimg.1997.0291

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: characterizing differential responses. *NeuroImage*, *7*(1), 30–40. doi:10.1006/nimg.1997.0306

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage*, *19*(4), 1273–302.

Friston, K. J., Josephs, O., Rees, G., & Turner, R. (1998). Nonlinear event-related responses in fMRI. *Magnetic Resonance in Medicine : Official Journal of the Society of Magnetic Resonance in Medicine*, *39*(1), 41–52.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(6), 3314–9.

Gauthier, I., Tarr, M. J., Moylan, J., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). The fusiform "face area" is part of a network that processes faces at the individual level. *Journal of Cognitive Neuroscience*, *12*(3), 495–504.

Gebhard, J. W., & Mowbray, G. H. (1959). On discriminating the rate of visual flicker and auditory flutter. *The American Journal of Psychology*, *72*, 521–9.

Giani, A. S., Ortiz, E., Belardinelli, P., Kleiner, M., Preissl, H., & Noppeney, U. (2012). Steady-state responses in MEG demonstrate information integration within but not across the auditory and visual senses. *NeuroImage*, *60*(2), 1478–89. doi:10.1016/j.neuroimage.2012.01.114

Giard, M. H., & Peronnet, F. (1999). Auditory-Visual Integration during Multimodal Object Recognition in Humans: A Behavioral and Electrophysiological Study. *Journal of Cognitive Neuroscience*, *11*(5), 473–490. doi:10.1162/089892999563544

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279–325.

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar Timing in Memory. *Annals of the New York Academy of Sciences*, *423*(1 Timing and Ti), 52–77. doi:10.1111/j.1749-6632.1984.tb23417.x

Glickstein, M., & Whitteridge, D. (1987). Tatsuji Inouye and the mapping of the visual fields on the human cerebral cortex. *Trends in Neurosciences*, *10*(9), 350–353. doi:10.1016/0166-2236(87)90066-X

Goense, J. B. M., & Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology : CB*, *18*(9), 631–40. doi:10.1016/j.cub.2008.03.054

Goffaux, V., & Rossion, B. (2007). Face inversion disproportionately impairs the perception of vertical but not horizontal relations between features. *Journal of Experimental Psychology. Human Perception and Performance*, *33*(4), 995–1002. doi:10.1037/0096-1523.33.4.995

Golla, H., Ignashchenkova, A., Haarmeier, T., & Thier, P. (2004). Improvement of visual acuity by spatial cueing: a comparative study in human and non-human primates. *Vision Research*, *44*(13), 1589–600. doi:10.1016/j.visres.2004.01.009

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. doi:10.1016/0166-2236(92)90344-8

Grill-Spector, K., Knouf, N., & Kanwisher, N. (2004). The fusiform face area subserves face perception, not generic within-category identification. *Nature Neuroscience*, *7*(5), 555–62. doi:10.1038/nn1224

Grondin, S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception & Psychophysics*, *72*(3), 561–82. doi:10.3758/APP.72.3.561

Gross, C. G. (1992). Representation of visual stimuli in inferior temporal cortex. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *335*(1273), 3–10. doi:10.1098/rstb.1992.0001

Gross, C. G., Rocha-Miranda, C. E., & Bender, D. B. (1972). Visual properties of neurons in inferotemporal cortex of the Macaque. *Journal of Neurophysiology*, *35*(1), 96–111.

Grueter, M., Grueter, T., Bell, V., Horst, J., Laskowski, W., Sperling, K., … Kennerknecht, I. (2007). Hereditary prosopagnosia: the first case series. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *43*(6), 734–49.

Grüter, T., Grüter, M., & Carbon, C.-C. (2008). Neural and genetic foundations of face recognition and prosopagnosia. *Journal of Neuropsychology*, *2*(Pt 1), 79–97.

Gurnsey, R., Pearson, P., & Day, D. (1996). Texture segmentation along the horizontal meridian: nonmonotonic changes in performance with eccentricity. *Journal of Experimental Psychology. Human Perception and Performance*, *22*(3), 738–57.

Haak, K. V, Winawer, J., Harvey, B. M., Renken, R., Dumoulin, S. O., Wandell, B. A., & Cornelissen, F. W. (2012). Connective field modeling. *NeuroImage*, *66C*, 376–384. doi:10.1016/j.neuroimage.2012.10.037

Hampel, F. R. (1974). The Influence Curve and its Role in Robust Estimation. *Journal of the American Statistical Association*, *69*(346), 383–393. doi:10.1080/01621459.1974.10482962

Handy, T. C., Soltani, M., & Mangun, G. R. (2001). Perceptual load and visuocortical processing: event-related potentials reveal sensory-level selection. *Psychological Science*, *12*(3), 213–8.

Harvey, B. M., & Dumoulin, S. O. (2011). The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(38), 13604–12. doi:10.1523/JNEUROSCI.2572-11.2011

Hasselmo, M. E., Rolls, E. T., & Baylis, G. C. (1989). The role of expression and identity in the face-selective responses of neurons in the temporal visual cortex of the monkey. *Behavioural Brain Research*, *32*(3), 203–218. doi:10.1016/S0166-4328(89)80054-3

Haxby, J., Hoffman, E., & Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*(6), 223–233.

Haynes, J.-D., & Rees, G. (2005a). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*(5), 686–91. doi:10.1038/nn1445

Haynes, J.-D., & Rees, G. (2005b). Predicting the stream of consciousness from activity in human visual cortex. *Current Biology : CB*, *15*(14), 1301–7. doi:10.1016/j.cub.2005.06.026

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews. Neuroscience*, *7*(7), 523–34. doi:10.1038/nrn1931

Heinemann, L., Kleinschmidt, A., & Müller, N. G. (2009). Exploring BOLD changes during spatial attention in non-stimulated visual cortex. *PloS One*, *4*(5), e5560. doi:10.1371/journal.pone.0005560

Hemond, C. C., Kanwisher, N. G., & Op de Beeck, H. P. (2007). A preference for contralateral stimuli in human object- and face-selective cortex. *PloS One*, *2*(6), e574. doi:10.1371/journal.pone.0000574

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory & Cognition*, *33*(1), 98–106.

Hills, P. J., Cooper, R. E., & Pake, J. M. (2013). First fixations in face processing: the more diagnostic they are the smaller the face-inversion effect. *Acta Psychologica*, *142*(2), 211–9. doi:10.1016/j.actpsy.2012.11.013

Hills, P. J., Ross, D. A., & Lewis, M. B. (2011). Attention misplaced: the role of diagnostic features in the face-inversion effect. *Journal of Experimental Psychology. Human Perception and Performance*, *37*(5), 1396–406. doi:10.1037/a0024247

Hills, P. J., Sullivan, A. J., & Pake, J. M. (2012). Aberrant first fixations when looking at inverted faces in various poses: the result of the centre-of-gravity effect? *British Journal of Psychology (London, England : 1953)*, *103*(4), 520–38. doi:10.1111/j.2044-8295.2011.02091.x

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–4. doi:10.1038/71152

Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., & Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *Journal of Computer Assisted Tomography*, *22*(2), 324–33.

Holmes, G. (1918). DISTURBANCES OF VISION BY CEREBRAL LESIONS. *The British Journal of Ophthalmology*, *2*(7), 353–84.

Holmes, N. P. (2009). The principle of inverse effectiveness in multisensory integration: some statistical considerations. *Brain Topography*, *21*(3-4), 168–76. doi:10.1007/s10548-009-0097-2

Hopf, J.-M., Boehler, C. N., Luck, S. J., Tsotsos, J. K., Heinze, H.-J., & Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(4), 1053–8. doi:10.1073/pnas.0507746103

Hopf, J.-M., Boehler, C. N., Schoenfeld, M. A., Heinze, H.-J., & Tsotsos, J. K. (2010). The spatial profile of the focus of attention in visual search: insights from MEG recordings. *Vision Research*, *50*(14), 1312–20. doi:10.1016/j.visres.2010.01.015

Hornak, J. P. (1996). The Basics of MRI. Retrieved August 21, 2013, from http://www.cis.rit.edu/htbooks/mri/

Horton, J. C., & Hoyt, W. F. (1991). The representation of the visual field in human striate cortex. A revision of the classic Holmes map. *Archives of Ophthalmology*, *109*(6), 816–24.

Hsiao, J. H., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science*, *19*(10), 998–1006. doi:10.1111/j.1467-9280.2008.02191.x

Hsieh, P.-J., Colas, J. T., & Kanwisher, N. (2012). Spatial pattern of BOLD fMRI activation reveals cross-modal information in auditory cortex. *Journal of Neurophysiology*, *107*(12), 3428–32. doi:10.1152/jn.01094.2010

Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, *13*(2), 415–25. doi:10.1109/72.991427

Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., & Turner, R. (2002). Image distortion correction in fMRI: A quantitative evaluation. *NeuroImage*, *16*(1), 217–40. doi:10.1006/nimg.2001.1054

Issa, E. B., & DiCarlo, J. J. (2012). Precedence of the eye region in neural processing of faces. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(47), 16666–82. doi:10.1523/JNEUROSCI.2391-12.2012

Ito, M., Tamura, H., Fujita, I., & Tanaka, K. (1995). Size and position invariance of neuronal responses in monkey inferotemporal cortex. *Journal of Neurophysiology*, *73*(1), 218–26.

Iurilli, G., Ghezzi, D., Olcese, U., Lassi, G., Nazzaro, C., Tonini, R., … Medini, P. (2012). Sound-driven synaptic inhibition in primary visual cortex. *Neuron*, *73*(4), 814–28. doi:10.1016/j.neuron.2011.12.026

Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, *12*(7), 273–80. doi:10.1016/j.tics.2008.04.002

James, T. W., Arcurio, L. R., & Gold, J. M. (2013). Inversion effects in face-selective cortex with combinations of face parts. *Journal of Cognitive Neuroscience*, *25*(3), 455–64. doi:10.1162/jocn_a_00312

James, W. (1890). *The principles of psychology, Vol 2*. *Natures Purpose* (1st ed., Vol. I, p. 722). New York: Henry Holt and Co.

Jemel, B., Mottron, L., & Dawson, M. (2006). Impaired face processing in autism: fact or artifact? *Journal of Autism and Developmental Disorders*, *36*(1), 91–106. doi:10.1007/s10803-005-0050-5

Johnston, A., Arnold, D. H., & Nishida, S. (2006). Spatially localized distortions of event time. *Current Biology : CB*, *16*(5), 472–9. doi:10.1016/j.cub.2006.01.032

Jones, M. R. (1976). Time, our lost dimension: toward a new theory of perception, attention, and memory. *Psychological Review*, *83*(5), 323–55.

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry*, *65*(8), 946–54. doi:10.1001/archpsyc.65.8.946

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–31. doi:10.1038/nature12715

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–85. doi:10.1038/nn1444

Kamke, M. R., Vieth, H. E., Cottrell, D., & Mattingley, J. B. (2012). Parietal disruption alters audiovisual binding in the sound-induced flash illusion. *NeuroImage*, *62*(3), 1334–1341. doi:10.1016/j.neuroimage.2012.05.063

Kanai, R., Dong, M. Y., Bahrami, B., & Rees, G. (2011). Distractibility in daily life is reflected in the structure and function of human parietal cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(18), 6620–6. doi:10.1523/JNEUROSCI.5864-10.2011

Kanai, R., & Rees, G. (2011). The structural basis of inter-individual differences in human behaviour and cognition. *Nature Reviews. Neuroscience*, *12*(4), 231–42. doi:10.1038/nrn3000

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(11), 4302–11.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: a cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *361*(1476), 2109–28. doi:10.1098/rstb.2006.1934

Karmarkar, U. R., & Buonomano, D. V. (2007). Timing in the absence of clocks: encoding time in neural network states. *Neuron*, *53*(3), 427–38. doi:10.1016/j.neuron.2007.01.006

Kasiński, A., Florek, A., & Schmidt, A. (2008). The put face database. *Image Processing & Communications*, *Vol. 13, n*, 59–64.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*(4), 751–61.

Kay, K. N., Naselaris, T., Prenger, R. J., & Gallant, J. L. (2008). Identifying natural images from human brain activity. *Nature*, *452*(7185), 352–5. doi:10.1038/nature06713

Kay, K. N., Winawer, J., Mezer, A., & Wandell, B. A. (2013). Compressive spatial summation in human visual cortex. *Journal of Neurophysiology*, *110*(2), 481–94. doi:10.1152/jn.00105.2013

Kayser, C., Logothetis, N. K., & Panzeri, S. (2010). Visual enhancement of the information representation in auditory cortex. *Current Biology : CB*, *20*(1), 19–24. doi:10.1016/j.cub.2009.10.068

Kehrer, L. (1987). Perceptual segregation and retinal position. *Spatial Vision*, *2*(4), 247–61.

Kehrer, L. (1989). Central performance drop on perceptual segregation tasks. *Spatial Vision*, *4*(1), 45–62.

Kehrer, L. (1997). The central performance drop in texture segmentation: a simulation based on a spatial filter model. *Biological Cybernetics*, *77*(4), 297–305. doi:10.1007/s004220050391

Kim, R., Peters, M. A. K., & Shams, L. (2011). 0 + 1 > 1: How Adding Noninformative Sound Improves Performance on a Visual Task.

Psychological Science, *32*(November 2011), 6–12. doi:10.1177/0956797611420662

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*(6), 14. doi:10.1016/0028-3932(83)90075-1

Klemen, J., & Chambers, C. D. (2012). Current perspectives and methods in studying neural mechanisms of multisensory interactions. *Neuroscience and Biobehavioral Reviews*, *36*(1), 111–33. doi:10.1016/j.neubiorev.2011.04.015

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*(9), 809–16.

Klin, A., Sparrow, S. S., de Bildt, A., Cicchetti, D. V, Cohen, D. J., & Volkmar, F. R. (1999). A normed study of face recognition in autism and related disorders. *Journal of Autism and Developmental Disorders*, *29*(6), 499–508.

Klink, P. C., Montijn, J. S., & Van Wezel, R. J. A. (2011). Crossmodal duration perception involves perceptual grouping, temporal ventriloquism, and variable internal clock rates. *Attention Perception Psychophysics*, *73*(1), 219–236.

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PloS One*, *2*(9), e943. doi:10.1371/journal.pone.0000943

Kovács, G., Cziraki, C., Vidnyánszky, Z., Schweinberger, S. R., & Greenlee, M. W. (2008). Position-specific and position-invariant face aftereffects reflect the adaptation of different cortical areas. *NeuroImage*, *43*(1), 156–64. doi:10.1016/j.neuroimage.2008.06.042

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, *17*(1), 26–49. doi:10.1016/j.tics.2012.10.011

Kravitz, D. J., Vinson, L. D., & Baker, C. I. (2008). How position dependent is visual object recognition? *Trends in Cognitive Sciences*, *12*(3), 114–22. doi:10.1016/j.tics.2007.12.006

Kriegeskorte, N. (2009). Relating Population-Code Representations between Man, Monkey, and Computational Models. *Frontiers in Neuroscience*, *3*(3), 363–73. doi:10.3389/neuro.01.035.2009

Kriegeskorte, N., Formisano, E., Sorger, B., & Goebel, R. (2007). Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(51), 20600–5. doi:10.1073/pnas.0705654104

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–8. doi:10.1073/pnas.0600244103

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–12. doi:10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., & Henriksson, L. (2013). Faciotopy - a face-feature map with face-like topology in the occipital face area. *Journal of Vision*, *13*(9), 1112–1112. doi:10.1167/13.9.1112

Lagarias, J. C., Reeds, J. A., Wright, M. H., & Wright, P. E. (1998). Convergence Properties of the Nelder--Mead Simplex Method in Low Dimensions. *SIAM Journal on Optimization*, *9*(1), 112–147. doi:10.1137/S1052623496303470

Laguesse, R., & Rossion, B. (2013). Face perception is whole or none: disentangling the role of spatial contiguity and interfeature distances in the composite face illusion. *Perception*, *42*(10), 1013–26.

Lakatos, P., Chen, C., Connell, M. N. O., Mills, A., & Schroeder, C. E. (2007). Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex. *Neuron*, 279–292. doi:10.1016/j.neuron.2006.12.011

Lakatos, P., O'Connell, M. N., Barczak, A., Mills, A., Javitt, D. C., & Schroeder, C. E. (2009). The leading sense: supramodal control of neurophysiological context by attention. *Neuron*, *64*(3), 419–430.

Large, E., & Jones, M. (1999). The Dynamics of Attending: How People Track Time-Varying Events. *Psychological Review*, *106*(1), 119 – 159.

Larsson, J., & Heeger, D. J. (2006). Two retinotopic visual areas in human lateral occipital cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *26*(51), 13128–42. doi:10.1523/JNEUROSCI.1657-06.2006

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology. Human Perception and Performance*, *21*(3), 451–68.

Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends in Cognitive Sciences*, *9*(2), 75–82. doi:10.1016/j.tics.2004.12.004

Lavie, N. (2010). Attention, Distraction, and Cognitive Control Under Load. *Current Directions in Psychological Science*, *19*(3), 143–148. doi:10.1177/0963721410370295

Lavie, N., & Cox, S. (1997). On the Efficiency of Visual Selective Attention: Efficient Visual Search Leads to Inefficient Distractor Rejection. *Psychological Science*, *8*(5), 395–396. doi:10.1111/j.1467-9280.1997.tb00432.x

Lavie, N., & Fox, E. (2000). The role of perceptual load in negative priming. *Journal of Experimental Psychology. Human Perception and Performance*, *26*(3), 1038–52.

Lavie, N., Hirst, A., de Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology. General*, *133*(3), 339–54. doi:10.1037/0096-3445.133.3.339

Lavie, N., Lin, Z., Zokaei, N., & Thoma, V. (2009). The role of perceptual load in object recognition. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(5), 1346–58. doi:10.1037/a0016454

Lavie, N., & Tsal, Y. (1994). Perceptual load as a major determinant of the locus of selection in visual attention. *Perception & Psychophysics*, *56*(2), 183–97.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: the role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *53*(2), 513–36. doi:10.1080/713755889

Lee, J. H., Durand, R., Gradinaru, V., Zhang, F., Goshen, I., Kim, D.-S., … Deisseroth, K. (2010). Global and local fMRI signals driven by neurons defined optogenetically by type and wiring. *Nature*, *465*(7299), 788–92. doi:10.1038/nature09108

Lee, S., Papanikolaou, A., Logothetis, N. K., Smirnakis, S. M., & Keliris, G. A. (2013). A new method for estimating population receptive field topography in visual cortex. *NeuroImage*, *81*, 144–57. doi:10.1016/j.neuroimage.2013.05.026

Leo, F., Romei, V., Freeman, E., Ladavas, E., & Driver, J. (2011). Looming sounds enhance orientation sensitivity for visual stimuli on the same side as such sounds. *Experimental Brain Research*, *213*(2-3), 193–201.

Leopold, D. A., Bondar, I. V, & Giese, M. A. (2006). Norm-based face encoding by single neurons in the monkey inferotemporal cortex. *Nature*, *442*(7102), 572–5. doi:10.1038/nature04951

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced shape encoding revealed by high-level aftereffects. *Nature Neuroscience*, *4*(1), 89–94. doi:10.1038/82947

Levin, N., Dumoulin, S. O., Winawer, J., Dougherty, R. F., & Wandell, B. A. (2010). Cortical maps and white matter tracts following long period of visual deprivation and retinal image restoration. *Neuron*, *65*(1), 21–31. doi:10.1016/j.neuron.2009.12.006

Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: an FMRI study. *Journal of Cognitive Neuroscience*, *22*(1), 203–11. doi:10.1162/jocn.2009.21203

Liu, T., Pestilli, F., & Carrasco, M. (2005). Transient attention enhances perceptual performance and FMRI response in human visual cortex. *Neuron*, *45*(3), 469–77. doi:10.1016/j.neuron.2004.12.039

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–78. doi:10.1038/nature06976

Logothetis, N. K. (2010). Bold claims for optogenetics. *Nature*, *468*(7323), E3–4; discussion E4–5. doi:10.1038/nature09532

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, a. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*(6843), 150–7. doi:10.1038/35084005

Logothetis, N. K., Pauls, J., & Poggio, T. (1995). Shape representation in the inferior temporal cortex of monkeys. *Current Biology : CB*, *5*(5), 552–63.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, *77*(1), 24–42.

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory Cortex Tracks Both Auditory and Visual Stimulus Dynamics Using Low-Frequency Neuronal Phase Modulation. *PLoS Biology*, *8*(8), e1000445. doi:10.1371/journal.pbio.1000445

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science (New York, N.Y.)*, *289*(5482), 1206–8.

Macmillan, N. A., & Creelman, C. D. (1997). d'plus: A program to calculate accuracy and bias measures from detection and discrimination data. *Spatial Vision*, *11*(1), 141–143.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide* (2nd Editio., p. 512). Psychology Press.

Magri, C., Schridde, U., Murayama, Y., Panzeri, S., & Logothetis, N. K. (2012). The amplitude and timing of the BOLD signal reflects the relationship between local field potential power at different frequencies. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(4), 1395–407. doi:10.1523/JNEUROSCI.3985-11.2012

Martínez, A., Anllo-Vento, L., Sereno, M. I., Frank, L. R., Buxton, R. B., Dubowitz, D. J., … Hillyard, S. A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. *Nature Neuroscience*, *2*(4), 364–9. doi:10.1038/7274

Martuzzi, R., Murray, M. M., Michel, C. M., Thiran, J.-P., Maeder, P. P., Clarke, S., & Meuli, R. A. (2007). Multisensory interactions within human primary cortices revealed by BOLD dynamics. *Cerebral Cortex*, *17*(7), 1672–9. doi:10.1093/cercor/bhl077

Maurer, D., Grand, R. Le, & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260. doi:10.1016/S1364-6613(02)01903-4

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–8.

McPartland, J. C., Webb, S. J., Keehn, B., & Dawson, G. (2011). Patterns of visual attention to faces and objects in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *41*(2), 148–57. doi:10.1007/s10803-010-1033-8

McPartland, J., Dawson, G., Webb, S. J., Panagiotides, H., & Carver, L. J. (2004). Event-related brain potentials reveal anomalies in temporal processing of faces in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *45*(7), 1235–45. doi:10.1111/j.1469-7610.2004.00318.x

McRobbie, D. W., Moore, E. A., Graves, M. J., & Prince, M. R. (2007). *MRI from Picture to Proton* (p. 394). Cambridge University Press.

McWeeny, K. H., Young, A. W., Hay, D. C., & Ellis, A. W. (1987). Putting names to faces. *British Journal of Psychology*, *78*(2), 143–149. doi:10.1111/j.2044-8295.1987.tb02235.x

Meienbrock, A., Naumer, M. J., Doehrmann, O., Singer, W., & Muckli, L. (2007). Retinotopic effects during spatial audio-visual integration. *Neuropsychologia*, *45*(3), 531–9. doi:10.1016/j.neuropsychologia.2006.05.018

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *7*(10), 3215–29.

Mesulam, M. M. (1998). From sensation to cognition. *Brain : A Journal of Neurology*, *121 (Pt 6*, 1013–52.

Meyer, K., Kaplan, J. T., Essex, R., Webber, C., Damasio, H., & Damasio, A. (2010). Predicting visual stimuli on the basis of activity in auditory cortices. *Nature Neuroscience*, *13*(6), 667–8. doi:10.1038/nn.2533

Miellet, S., Vizioli, L., He, L., Zhou, X., & Caldara, R. (2013). Mapping Face Recognition Information Use across Cultures. *Frontiers in Psychology*, *4*, 34. doi:10.3389/fpsyg.2013.00034

Misaki, M., Kim, Y., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage*, *53*(1), 103–18. doi:10.1016/j.neuroimage.2010.05.051

Mishra, J., Martínez, A., & Hillyard, S. a. (2010). Effect of attention on early cortical processes associated with the sound-induced extra flash illusion. *Journal of Cognitive Neuroscience*, *22*(8), 1714–29. doi:10.1162/jocn.2009.21295

Mishra, J., Martinez, A., Sejnowski, T. J., & Hillyard, S. a. (2007). Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(15), 4120–31. doi:10.1523/JNEUROSCI.4912-06.2007

Molholm, S., Ritter, W., Murray, M. M., Javitt, D. C., Schroeder, C. E., & Foxe, J. J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Research. Cognitive Brain Research*, *14*(1), 115–28.

Montagna, B., Pestilli, F., & Carrasco, M. (2009). Attention trades off spatial acuity. *Vision Research*, *49*(7), 735–45.

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science (New York, N.Y.)*, *229*(4715), 782–4.

Moraschi, M., DiNuzzo, M., & Giove, F. (2012). On the origin of sustained negative BOLD response. *Journal of Neurophysiology*, *108*(9), 2339–42. doi:10.1152/jn.01199.2011

Morawetz, C., Holz, P., Baudewig, J., Treue, S., & Dechent, P. (2007). Split of attentional resources in human visual cortex. *Visual Neuroscience*, *24*(6), 817–26. doi:10.1017/S0952523807070745

Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: examining temporal ventriloquism. *Brain Research. Cognitive Brain Research*, *17*(1), 154–63.

Morey, R. D. (2008). Confidence Intervals from Normalized Data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64.

Morgan, M. (2013). Sustained attention is not necessary for velocity adaptation. *Journal of Vision*, *13*(8), 26–. doi:10.1167/13.8.26

Morgan, M. J. (2011). Wohlgemuth was right: distracting attention from the adapting stimulus does not decrease the motion after-effect. *Vision Research*, *51*(20), 2169–75. doi:10.1016/j.visres.2011.07.018

Morgan, M. J. (2012). Motion adaptation does not depend on attention to the adaptor. *Vision Research*, *55*, 47–51.

Morgan, M. L., Deangelis, G. C., & Angelaki, D. E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, *59*(4), 662–73. doi:10.1016/j.neuron.2008.06.024

Muggleton, N., Lamb, R., Walsh, V., & Lavie, N. (2008). Perceptual load modulates visual cortex excitability to magnetic stimulation. *Journal of Neurophysiology*, *100*(1), 516–9. doi:10.1152/jn.01287.2007

Muller, N. G., Bartelt, O. A., Donner, T. H., Villringer, A., & Brandt, S. A. (2003). A Physiological Correlate of the "Zoom Lens" of Visual Attention. *J. Neurosci.*, *23*(9), 3561–3565.

Müller, N. G., & Kleinschmidt, A. (2004). The attentional "spotlight"s' penumbra: center-surround modulation in striate cortex. *Neuroreport*, *15*(6), 977–80.

Murray, M. M., Cappe, C., Romei, V., Martuzzi, R., & Thut, G. (2012). Auditory-visual multisensory interactions in humans: synthesis and controversies. In B. Stein (Ed.), *The New Handbook of Multisensory Processing*. MIT press.

Nath, A. R., & Beauchamp, M. S. (2012). A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. *NeuroImage*, *59*(1), 781–7. doi:10.1016/j.neuroimage.2011.07.024

Nath, A. R., Fava, E. E., & Beauchamp, M. S. (2011). Neural correlates of interindividual differences in children's audiovisual speech perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(39), 13963–71. doi:10.1523/JNEUROSCI.2605-11.2011

Naue, N., Rach, S., Strüber, D., Huster, R. J., Zaehle, T., Körner, U., & Herrmann, C. S. (2011). Auditory event-related response in visual cortex modulates subsequent visual responses in humans. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(21), 7729–36. doi:10.1523/JNEUROSCI.1076-11.2011

Nelder, J. A., & Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, *7*(4), 308–313. doi:10.1093/comjnl/7.4.308

Nichols, D. F., Betts, L. R., & Wilson, H. R. (2010). Decoding of faces and face components in face-sensitive human visual cortex. *Frontiers in Psychology*, *1*, 28. doi:10.3389/fpsyg.2010.00028

Niebergall, R., Khayat, P. S., Treue, S., & Martinez-Trujillo, J. C. (2011). Multifocal attention filters targets from distracters within and beyond primate MT neurons' receptive field boundaries. *Neuron*, *72*(6), 1067–79. doi:10.1016/j.neuron.2011.10.013

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. *Current Biology : CB*, *21*(19), 1641–6. doi:10.1016/j.cub.2011.08.031

Noesselt, T., Rieger, J. W., Schoenfeld, M. A., Kanowski, M., Hinrichs, H., Heinze, H.-J., & Driver, J. (2007). Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(42), 11431–41. doi:10.1523/JNEUROSCI.2252-07.2007

Noesselt, T., Tyll, S., Boehler, C. N., Budinger, E., Heinze, H.-J., & Driver, J. (2010). Sound-Induced Enhancement of Low-Intensity Vision: Multisensory Influences on Human Sensory-Specific Cortices and Thalamic Bodies Relate to Perceptual Enhancement of Visual Detection Sensitivity. *Journal of Neuroscience*, *30*(41), 13609–13623. doi:10.1523/JNEUROSCI.4524-09.2010

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–30. doi:10.1016/j.tics.2006.07.005

O'Connell, R. G., Schneider, D., Hester, R., Mattingley, J. B., & Bellgrove, M. A. (2011). Attentional load asymmetrically affects early electrophysiological indices of visual orienting. *Cerebral Cortex (New York, N.Y. : 1991)*, *21*(5), 1056–65. doi:10.1093/cercor/bhq178

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, *5*(11), 1203–9. doi:10.1038/nn957

Ohta, H., Yamada, T., Watanabe, H., Kanai, C., Tanaka, E., Ohno, T., … Hashimoto, R.-I. (2012). An fMRI study of reduced perceptual load-dependent modulation of task-irrelevant activity in adults with autism spectrum conditions. *NeuroImage*, *61*(4), 1176–87. doi:10.1016/j.neuroimage.2012.03.042

Op De Beeck, H., & Vogels, R. (2000). Spatial sensitivity of macaque inferior temporal neurons. *The Journal of Comparative Neurology*, *426*(4), 505–18.

Oppelt, A. (1983). Kernmagnetische Resonanz in der Medizin. *Physik in Unserer Zeit*, *14*(1), 7–17. doi:10.1002/piuz.19830140102

Osterling, J. A., Dawson, G., & Munson, J. A. (2002). Early recognition of 1-year-old infants with autism spectrum disorder versus mental retardation. *Development and Psychopathology*, *14*(2), 239–51.

Parvizi, J., Jacques, C., Foster, B. L., Witthoft, N., Withoft, N., Rangarajan, V., … Grill-Spector, K. (2012). Electrical stimulation of human fusiform face-selective regions distorts face perception. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *32*(43), 14915–20. doi:10.1523/JNEUROSCI.2609-12.2012

Pascual-Leone, A., & Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science (New York, N.Y.)*, *292*(5516), 510–2. doi:10.1126/science.1057099

Pauling, L., & Coryell, C. D. (1936). The Magnetic Properties and Structure of Hemoglobin, Oxyhemoglobin and Carbonmonoxyhemoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, *22*(4), 210–6.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–42.

Pelphrey, K. A., Sasson, N. J., Reznick, J. S., Paul, G., Goldman, B. D., & Piven, J. (2002). Visual scanning of faces in autism. *Journal of Autism and Developmental Disorders*, *32*(4), 249–61.

Perret, D. I., Harries, M. H., Mistlin, A. J., Hietanen, J. K., Benson, P. J., BEVAN, R., … Brierley, K. (1990). Social signals analyzed at the single cell level : someone is looking at me, something touched me, something moved! *International Journal of Comparative Psychology*, *4*(1), 25–55.

Perrett, D. I., Rolls, E. T., & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, *47*(3). doi:10.1007/BF00239352

Perrett, D. I., Smith, P. A. J., Potter, D. D., Mistlin, A. J., Head, A. S., Milner, A. D., & Jeeves, M. A. (1985). Visual Cells in the Temporal Cortex Sensitive to Face View and Gaze Direction. *Proceedings of the Royal Society B: Biological Sciences*, *223*(1232), 293–317. doi:10.1098/rspb.1985.0003

Peterson, M. F., & Eckstein, M. P. (2012). Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(48), E3314–23. doi:10.1073/pnas.1214269109

Peterson, M. F., & Eckstein, M. P. (2013a). Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychological Science*, *24*(7), 1216–25. doi:10.1177/0956797612471684

Peterson, M. F., & Eckstein, M. P. (2013b). Learning optimal eye movements to unusual faces. *Vision Research*. doi:10.1016/j.visres.2013.11.005

Pfeuffer, J., McCullough, J. C., Van de Moortele, P. F., Ugurbil, K., & Hu, X. (2003). Spatial dependence of the nonlinear BOLD response at short stimulus duration. *NeuroImage*, *18*(4), 990–1000.

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception & Psychophysics*, *6*(4), 203–205. doi:10.3758/BF03207017

Pinsk, M. A., Doniger, G. M., & Kastner, S. (2004). Push-pull mechanism of selective attention in human extrastriate cortex. *Journal of Neurophysiology*, *92*(1), 622–9. doi:10.1152/jn.00974.2003

Pitcher, D., Duchaine, B., Walsh, V., Yovel, G., & Kanwisher, N. (2011). The role of lateral occipital face and object areas in the face inversion effect. *Neuropsychologia*, *49*(12), 3448–53. doi:10.1016/j.neuropsychologia.2011.08.020

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology : CB*, *17*(18), 1568–73. doi:10.1016/j.cub.2007.07.063

Plainis, S., Murray, I. J., & Chauhan, K. (2001). Raised visual detection thresholds depend on the level of complexity of cognitive foveal loading. *Perception*, *30*(10), 1203–12.

Poldrack, R. A., Mumford, J. A., & Nichols, T. E. (2011). *Handbook of Functional MRI Data Analysis* (p. 228). Cambridge University Press.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, *83*(2), 157–171. doi:10.1037/0033-295X.83.2.157

Potechin, C., & Gurnsey, R. (2003). Backward masking is not required to elicit the central performance drop. *Spatial Vision*, *16*(5), 393–406.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *16*(16), 5205–15.

Ramkumar, P., Jas, M., Pannasch, S., Hari, R., & Parkkonen, L. (2013). Feature-specific information processing precedes concerted activation in human visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(18), 7691–9. doi:10.1523/JNEUROSCI.3905-12.2013

Rauss, K., Pourtois, G., Vuilleumier, P., & Schwartz, S. (2012). Effects of attentional load on early visual processing depend on stimulus timing. *Human Brain Mapping*, *33*(1), 63–74. doi:10.1002/hbm.21193

Rauss, K. S., Pourtois, G., Vuilleumier, P., & Schwartz, S. (2009). Attentional load modifies early activity in human primary visual cortex. *Human Brain Mapping*, *30*(5), 1723–33. doi:10.1002/hbm.20636

Rees, G., Frith, C. D., & Lavie, N. (1997). Modulating irrelevant motion perception by varying attentional load in an unrelated task. *Science (New York, N.Y.)*, *278*(5343), 1616–9.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4. *J. Neurosci.*, *19*(5), 1736–1753.

Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, *61*(2), 168–85. doi:10.1016/j.neuron.2009.01.002

Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, *26*(3), 703–14.

Rhodes, G. (1988). Looking at faces: first-order and second-order features as determinants of facial appearance. *Perception*, *17*(1), 43–63.

Rhodes, G., & Jeffery, L. (2006). Adaptive norm-based coding of facial identity. *Vision Research*, *46*(18), 2977–87. doi:10.1016/j.visres.2006.03.002

Rhodes, G., Jeffery, L., Watson, T. L., Clifford, C. W. G., & Nakayama, K. (2003). Fitting the mind to the world: face adaptation and attractiveness aftereffects. *Psychological Science*, *14*(6), 558–66.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, *2*(11), 1019–25. doi:10.1038/14819

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *50*(1-2), 19–26.

Romei, V., De Haas, B., Mok, R. M., & Driver, J. (2011). Auditory Stimulus Timing Influences Perceived duration of Co-Occurring Visual Stimuli. *Frontiers in Psychology*, *2*(September), 8.

Romei, V., Gross, J., & Thut, G. (2012). Sounds reset rhythms of visual cortex and corresponding human visual perception. *Current Biology*, *22*(9), 807–813.

Romei, V., Murray, M. M., Cappe, C., & Thut, G. (2009). Preperceptual and stimulus-selective enhancement of low-level human visual cortex excitability by sounds. *Current Biology : CB*, *19*(21), 1799–805. doi:10.1016/j.cub.2009.09.027

Romei, V., Murray, M. M., Merabet, L. B., & Thut, G. (2007). Occipital transcranial magnetic stimulation has opposing effects on visual and auditory stimulus detection: implications for multisensory interactions. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *27*(43), 11465–72. doi:10.1523/JNEUROSCI.2827-07.2007

Rorden, C., Guerrini, C., Swainson, R., Lazzeri, M., & Baylis, G. C. (2008). Event related potentials reveal that increasing perceptual load leads to increased responses for target stimuli and decreased responses for irrelevant stimuli. *Frontiers in Human Neuroscience*, *2*, 4. doi:10.3389/neuro.09.004.2008

Rose, F. E., Lincoln, A. J., Lai, Z., Ene, M., Searcy, Y. M., & Bellugi, U. (2007). Orientation and affective expression effects on face recognition in Williams syndrome and autism. *Journal of Autism and Developmental Disorders*, *37*(3), 513–22. doi:10.1007/s10803-006-0200-4

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition*, *21*(2), 139–253. doi:10.1080/13506285.2013.772929

Russell, C., Malhotra, P., & Husain, M. (2004). Attention modulates the visual field in healthy observers and parietal patients. *Neuroreport*, *15*(14), 2189–93.

Sato, Y., Toyoizumi, T., & Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: identification of common sources of audiovisual stimuli. *Neural Computation*, *19*(12), 3335–55. doi:10.1162/neco.2007.19.12.3335

Schira, M. M., Tyler, C. W., Breakspear, M., & Spehar, B. (2009). The foveal confluence in human visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *29*(28), 9050–8. doi:10.1523/JNEUROSCI.1760-09.2009

Schmalzl, L., Palermo, R., Green, M., Brunsdon, R., & Coltheart, M. (2008). Training of familiar face recognition and visual scan paths for faces in a child with congenital prosopagnosia. *Cognitive Neuropsychology*, *25*(5), 704–29. doi:10.1080/02643290802299350

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences*, *12*(3), 106–13. doi:10.1016/j.tics.2008.01.002

Schwaninger, A., Carbon, C. C., & Leder, H. (2003). Expert face processing: Specialization and constraints of face processing. In G. Schwarzer & H. Leder (Eds.), *Development of face processing* (pp. 81–97).

Schwaninger, A., & Mast, F. W. (2005). The face-inversion effect can be explained by the capacity limitations of an orientation normalization mechanism1. *Japanese Psychological Research*, *47*(3), 216–222. doi:10.1111/j.1468-5884.2005.00290.x

Schwartz, S., Vuilleumier, P., Hutton, C., Maravita, A., Dolan, R. J., & 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. *Cerebral Cortex (New York, N.Y. : 1991)*, *15*(6), 770–86. doi:10.1093/cercor/bhh178

Schwarzer, G., Huber, S., Grüter, M., Grüter, T., Gross, C., Hipfel, M., & Kennerknecht, I. (2007). Gaze behaviour in hereditary prosopagnosia.

Psychological Research, *71*(5), 583–90. doi:10.1007/s00426-006-0068-0

Schwarzkopf, D. S., Song, C., & Rees, G. (2011). The surface area of human V1 predicts the subjective experience of object size. *Nature Neuroscience*, *14*(1), 28–30. doi:10.1038/nn.2706

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(11), 4447–52. doi:10.1073/pnas.0800431105

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., … Tootell, R. B. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science (New York, N.Y.)*, *268*(5212), 889–93.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(15), 6424–9. doi:10.1073/pnas.0700622104

Shalev, L., & Tsal, Y. (2002). Detecting gaps with and without attention: Further evidence for attentional receptive fields. *European Journal of Cognitive Psychology*, *14*(1), 3–26. doi:10.1080/09541440143000005

Shams, L., Iwaki, S., Chawla, A., & Bhattacharya, J. (2005). Early modulation of visual cortex by sound : an MEG study. *Neuroscience Letters*, *378*, 76–81. doi:10.1016/j.neulet.2004.12.035

Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions: What you see is what you hear. *Nature*, *408*(December), 788.

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain Research. Cognitive Brain Research*, *14*(1), 147–52.

Shams, L., Ma, W. J., & Beierholm, U. (2005). Sound-induced flash illusion as an optimal percept. *Neuroreport*, *16*(17), 1923–7.

Shipley, T. (1964). Auditory Flutter-Driving of Visual Flicker. *Science*, *145*(3638), 1328–1330.

Shiu, L. P., & Pashler, H. (1995). Spatial attention and vernier acuity. *Vision Research*, *35*(3), 337–43.

Shmuel, A., Augath, M., Oeltermann, A., & Logothetis, N. K. (2006). Negative functional MRI response correlates with decreases in neuronal activity

in monkey visual area V1. *Nature Neuroscience*, *9*(4), 569–77. doi:10.1038/nn1675

Silver, M. A., Shenhav, A., & D'Esposito, M. (2008). Cholinergic enhancement reduces spatial spread of visual responses in human early visual cortex. *Neuron*, *60*(5), 904–14. doi:10.1016/j.neuron.2008.09.038

Sirotin, Y. B., & Das, A. (2009). Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. *Nature*, *457*(7228), 475–9. doi:10.1038/nature07664

Smith, M. L., Fries, P., Gosselin, F., Goebel, R., & Schyns, P. G. (2009). Inverse mapping the neuronal substrates of face categorizations. *Cerebral Cortex (New York, N.Y. : 1991)*, *19*(10), 2428–38. doi:10.1093/cercor/bhn257

Smith, M. L., Gosselin, F., & Schyns, P. G. (2004). Receptive fields for flexible face categorizations. *Psychological Science*, *15*(11), 753–61. doi:10.1111/j.0956-7976.2004.00752.x

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology. General*, *117*(1), 34–50.

Spence, C., & Driver, J. (2004). *Crossmodal space and crossmodal attention*. (C. Spence & J. Driver, Eds.)*Journal of Psychophysiology* (Vol. 19, pp. 141–177). OUP.

Spence, C., & Squire, S. (2003). Multisensory integration: maintaining the perception of synchrony. *Current Biology : CB*, *13*(13), R519–21.

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science (New York, N.Y.)*, *240*(4850), 338–40.

Staeren, N., Renvall, H., De Martino, F., Goebel, R., & Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Current Biology : CB*, *19*(6), 498–502. doi:10.1016/j.cub.2009.01.066

Stelzer, J., Chen, Y., & Turner, R. (2013). Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *NeuroImage*, *65*, 69–82. doi:10.1016/j.neuroimage.2012.09.063

Stephan, B. C. M., & Caine, D. (2009). Aberrant pattern of scanning in prosopagnosia reflects impaired face processing. *Brain and Cognition*, *69*(2), 262–8. doi:10.1016/j.bandc.2008.07.015

Stone, J. V, Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, M., … Porter, N. R. (2001). When is now? Perception of simultaneity. *Proceedings. Biological Sciences / The Royal Society*, *268*(1462), 31–8. doi:10.1098/rspb.2000.1326

Susilo, T., Rezlescu, C., & Duchaine, B. (2013). The composite effect for inverted faces is reliable at large sample sizes and requires the basic face configuration. *Journal of Vision*, *13*(13), 14. doi:10.1167/13.13.14

Talgar, C. P., & Carrasco, M. (2002). Vertical meridian asymmetry in spatial resolution: visual and attentional factors. *Psychonomic Bulletin & Review*, *9*(4), 714–22.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *46*(2), 225–245. doi:10.1080/14640749308401045

Tanaka, J. W., & Sung, A. (2013). The "Eye Avoidance" Hypothesis of Autism Face Processing. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-013-1976-7

Thorne, J. D., De Vos, M., Viola, F. C., & Debener, S. (2011). Cross-modal phase reset predicts auditory task performance in humans. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(10), 3853–61. doi:10.1523/JNEUROSCI.6176-10.2011

Tootell, R. B., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., & Dale, A. M. (1998). The retinotopy of visual spatial attention. *Neuron*, *21*(6), 1409–22.

Tovee, M. J., Rolls, E. T., & Azzopardi, P. (1994). Translation invariance in the responses to faces of single neurons in the temporal visual cortical areas of the alert macaque. *J Neurophysiol*, *72*(3), 1049–1060.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review*, *76*(3), 282–99.

Tsao, D. Y., Freiwald, W. A., Knutsen, T. A., Mandeville, J. B., & Tootell, R. B. H. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, *6*(9), 989–95. doi:10.1038/nn1111

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., & Livingstone, M. S. (2006). A cortical region consisting entirely of face-selective cells. *Science (New York, N.Y.)*, *311*(5761), 670–4. doi:10.1126/science.1119983

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. *Trends in Cognitive Sciences*, *11*(2), 58–64. doi:10.1016/j.tics.2006.11.009

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology*, *79*, 471–491.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161–204. doi:10.1080/14640749108400966

Van Belle, G., Ramon, M., Lefèvre, P., & Rossion, B. (2010). Fixation patterns during recognition of personally familiar and unfamiliar faces. *Frontiers in Psychology*, *1*, 20. doi:10.3389/fpsyg.2010.00020

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science (New York, N.Y.)*, *255*(5043), 419–23.

Vazquez, A. L., & Noll, D. C. (1998). Nonlinear aspects of the BOLD response in functional MRI. *NeuroImage*, *7*(2), 108–18. doi:10.1006/nimg.1997.0316

Von Der Heide, R. J., Skipper, L. M., & Olson, I. R. (2013). Anterior temporal face patches: a meta-analysis and empirical study. *Frontiers in Human Neuroscience*, *7*, 17. doi:10.3389/fnhum.2013.00017

Von Saldern, S., & Noppeney, U. (2013). Sensory and striatal areas integrate auditory and visual signals into behavioral benefits during motion discrimination. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(20), 8841–9. doi:10.1523/JNEUROSCI.3020-12.2013

Vroomen, J., & De Gelder, B. (2000). Sound enhances visual perception: cross-modal effects of auditory organization on vision. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1583–1590.

Vroomen, J., & Keetels, M. (2009). Sounds change four-dot masking. *Acta Psychologica*, *130*(1), 58–63. doi:10.1016/j.actpsy.2008.10.001

Vroomen, J., & Keetels, M. (2010). Perception of intersensory synchrony: a tutorial review. *Attention, Perception & Psychophysics*, *72*(4), 871–84. doi:10.3758/APP.72.4.871

Vuilleumier, P., Schwartz, S., Verdon, V., Maravita, A., Hutton, C., Husain, M., & Driver, J. (2008). Abnormal attentional modulation of retinotopic cortex in parietal patients with spatial neglect. *Current Biology : CB*, *18*(19), 1525–9. doi:10.1016/j.cub.2008.08.072

Wada, Y., Kitagawa, N., & Noguchi, K. (2003). Audio-visual integration in temporal perception. *International Journal of Psychophysiology : Official*

Journal of the International Organization of Psychophysiology, *50*(1-2), 117–24.

Walker, J. T., & Scott, K. J. (1981). Auditory-visual conflicts in the perceived duration of lights, tones and gaps. *Journal of Experimental Psychology. Human Perception and Performance*, *7*(6), 1327–39.

Walker-Smith, G. J., Gale, A. G., & Findlay, J. M. (1977). Eye movement strategies involved in face perception. *Perception*, *6*(3), 313–26.

Wandell, B. a, Dumoulin, S. O., & Brewer, A. a. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366–83. doi:10.1016/j.neuron.2007.10.012

Wandell, B. A. (1995). *Foundations of vision* (p. 476). Sinauer Associates, Incorporated, Sunderland, Mass.

Wang, Y., Celebrini, S., Trotter, Y., & Barone, P. (2008). Visuo-auditory interactions in the primary visual cortex of the behaving monkey: electrophysiological evidence. *BMC Neuroscience*, *9*, 79. doi:10.1186/1471-2202-9-79

Watkins, S., Shams, L., Tanaka, S., Haynes, J.-D., & Rees, G. (2006). Sound alters activity in human V1 in association with illusory visual perception. *NeuroImage*, *31*(3), 1247–56. doi:10.1016/j.neuroimage.2006.01.016

Webb, B. S., Dhruv, N. T., Solomon, S. G., Tailby, C., & Lennie, P. (2005). Early and late mechanisms of surround suppression in striate cortex of macaque. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *25*(50), 11666–75. doi:10.1523/JNEUROSCI.3414-05.2005

Webster, M. A., Kaping, D., Mizokami, Y., & Duhamel, P. (2004). Adaptation to natural facial categories. *Nature*, *428*(6982), 557–61. doi:10.1038/nature02420

Webster, M. A., & MacLin, O. H. (1999). Figural aftereffects in the perception of faces. *Psychonomic Bulletin & Review*, *6*(4), 647–53.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: a review of behavioral studies. *Neuroscience and Biobehavioral Reviews*, *36*(3), 1060–84. doi:10.1016/j.neubiorev.2011.12.008

Weiner, K. S., Golarai, G., Caspers, J., Chuapoco, M. R., Mohlberg, H., Zilles, K., … Grill-Spector, K. (2014). The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human

ventral temporal cortex. *NeuroImage*, *84*, 453–65. doi:10.1016/j.neuroimage.2013.08.068

Weiner, K. S., & Grill-Spector, K. (2012). The improbable simplicity of the fusiform face area. *Trends in Cognitive Sciences*, *16*(5), 251–4. doi:10.1016/j.tics.2012.03.003

Weiner, K. S., & Grill-Spector, K. (2013). Neural representations of faces and limbs neighbor in human high-level visual cortex: evidence for a new organization principle. *Psychological Research*, *77*(1), 74–97. doi:10.1007/s00426-011-0392-x

Welch, R. B., DutionHurt, L. D., & Warren, D. H. (1986). Contributions of audition and vision to temporal rate perception. *Perception & Psychophysics*, *39*(4), 294–300. doi:10.3758/BF03204939

Werkhoven, P. J., van Erp, J. B. F., & Philippi, T. G. (2009). Counting visual and tactile events: the effect of attention on multisensory integration. *Attention, Perception & Psychophysics*, *71*(8), 1854–61. doi:10.3758/APP.71.8.1854

Werner, S., & Noppeney, U. (2010). Distinct functional contributions of primary sensory and association areas to audiovisual integration in object categorization. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *30*(7), 2662–75. doi:10.1523/JNEUROSCI.5091-09.2010

Willems, R. M., Peelen, M. V, & Hagoort, P. (2010). Cerebral lateralization of face-selective and body-selective visual areas depends on handedness. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(7), 1719–25. doi:10.1093/cercor/bhp234

Wilson, C. E., Palermo, R., & Brock, J. (2012). Visual scan paths and recognition of facial identity in autism spectrum disorder and typical development. *PloS One*, *7*(5), e37681. doi:10.1371/journal.pone.0037681

Winawer, J., Kay, K. N., Foster, B. L., Rauschecker, A. M., Parvizi, J., & Wandell, B. A. (2013). Asynchronous broadband signals are the principal source of the BOLD response in human visual cortex. *Current Biology : CB*, *23*(13), 1145–53. doi:10.1016/j.cub.2013.05.001

Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: merging auditory and visual worlds. *Neuron*, *48*(3), 489–96. doi:10.1016/j.neuron.2005.10.020

Womelsdorf, T., Anton-Erxleben, K., Pieper, F., & Treue, S. (2006). Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nature Neuroscience*, *9*(9), 1156–60. doi:10.1038/nn1748

Womelsdorf, T., Anton-Erxleben, K., & Treue, S. (2008). Receptive field shift and shrinkage in macaque middle temporal area through attentional gain modulation. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *28*(36), 8934–44. doi:10.1523/JNEUROSCI.4030-07.2008

Worsley, K. (2007). Random Field Theoory. In K. J. Friston, J. T. Ashburner, S. J. Kiebel, T. E. Nichols, & W. D. Penny (Eds.), *Statistical Parametric Mapping* (pp. 232–237). London: Academic Press.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature*, *396*(6706), 72–5. doi:10.1038/23936

Yeshurun, Y., & Carrasco, M. (1999). Spatial attention improves performance in spatial resolution tasks. *Vision Research*, *39*(2), 293–306.

Yeshurun, Y., & Carrasco, M. (2000). The locus of attentional effects in texture segmentation. *Nature Neuroscience*, *3*(6), 622–7. doi:10.1038/75804

Yeshurun, Y., Montagna, B., & Carrasco, M. (2008). On the flexibility of sustained attention and its effects on a texture segmentation task. *Vision Research*, *48*(1), 80–95. doi:10.1016/j.visres.2007.10.015

Yin, R. (1969). Looking at upside-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145.

Young, A. W., Hay, D. C., & Ellis, A. W. (1985). The faces that launched a thousand slips: everyday difficulties and errors in recognizing people. *British Journal of Psychology (London, England : 1953)*, *76 (Pt 4)*, 495–523.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configurational information in face perception. *Perception*, *16*(6), 747–59.

Yovel, G., & Kanwisher, N. (2005). The neural basis of the behavioral face-inversion effect. *Current Biology : CB*, *15*(24), 2256–62. doi:10.1016/j.cub.2005.10.072

Yue, X., Cassidy, B. S., Devaney, K. J., Holt, D. J., & Tootell, R. B. H. (2011). Lower-level stimulus features strongly influence responses in the fusiform face area. *Cerebral Cortex (New York, N.Y. : 1991)*, *21*(1), 35–47. doi:10.1093/cercor/bhq050

Yuille, A. L., & Buelthoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian Inference* (pp. 123–161). Cambridge University Press.

Zenger, B., Braun, J., & Koch, C. (2000). Attentional effects on contrast detection in the presence of surround masks. *Vision Research*, *40*(27), 3717–24.

Zimmer, M., & Kovács, G. (2011). Position specificity of adaptation-related face aftereffects. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*(1564), 586–95. doi:10.1098/rstb.2010.0265

Zuiderbaan, W., Harvey, B. M., & Dumoulin, S. O. (2012). Modeling center-surround configurations in population receptive fields using fMRI. *Journal of Vision*, *12*(3), 10. doi:10.1167/12.3.10