Bennett, Matthew (2018) *Investigating visual and auditory scene feedback to early visual foveal and peripheral cortex using fMRI.* PhD thesis.

https://theses.gla.ac.uk/38935/

Investigating Visual and Auditory Scene Feedback to Early Visual Foveal and Peripheral Cortex using fMRI

Matthew Bennett, BSc, MSc

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

Institute of Psychology & Neuroscience
College of Medical, Veterinary & Life Sciences
University of Glasgow

July 9th 2018

Contains studies performed in:
Centre for Cognitive Neuroimaging
Department of Psychology
University of Glasgow
Glasgow
G12 8QB
United Kingdom

Abstract

In this thesis I will present three lines of investigation using functional magnetic resonance imaging (fMRI) and multivariate pattern analysis (MVPA) methods to elucidate the characteristics of cortico-cortical feedback to early visual cortex. In particular, we were interested in whether regions retinotopically sensitive to central and peripheral regions of the visual field received different characteristic information types through feedback which suited their different processing capabilities.

In **Chapter 1**, I describe the organisation of the visual system in terms of its feed-forward and feedback processing streams. I also describe functional magnetic resonance imaging (fMRI) and the blood oxygenation level dependent (BOLD) signal and how it may be used to study visual feedback processes. I conclude with a brief rationale for studying feedback and in particular for why we focussed on feedback to foveal and peripheral early visual cortex.

In **Chapter 2**, we present three fMRI experiments that use the occluded paradigm to build on work showing that objects are fedback exclusively to foveal cortex. We also use a novel analysis technique to provide support for the idea that the feedback information into foveal V1 cortex originates, at least in part, from lateral occipital cortex.

In **Chapter 3**, we present two fMRI experiments in which we present subjects with 'hybrid images', which can be perceived as one of two natural scenes – one carried in low and one carried in high frequency information channels. We find that far peripheral early visual cortex is influenced in a top down manner, whilst foveal cortex more faithfully represents the feedforward input. This provides evidence for scene related feedback to far peripheral cortex.

In **Chapter 4**, we use high resolution 7T fMRI to study cross-modal feedback across the laminae of early visual cortex. We show that scene information presented in the auditory domain can be found in the deeper layers of early visual cortex. Conversely, mental scene imagery can be found in the superficial layers. We

conclude that high resolution 7T fMRI can shed new light on the interaction between feed forward and feedback with laminar resolution.

In **Chapter 5**, I briefly summarise and discuss the findings and draw conclusions. I point out some limitations to be addressed and make recommendations for future work to extend the results presented in this thesis.

Table of Contents

List of Figures

List of Publications

Oral Presentations

M A Bennett, L S Petro, L Muckli (2016): Investigating cortical feedback of background scene and embedded objects to foveal and peripheral V1 using fMRI. Vision Sciences Society, 16th Ann. Meeting, May 13-18, St. Pete Beach, Florida, USA.

Posters

M A Bennett, L S Petro, A. T. Morgan, F. De Martino, L Muckli (2017): Using high-resolution 7T fMRI reveals auditory and imagery information in non-stimulated visual cortex. Society for Neuroscience, Nov, 11 - 15, Washington DC, USA.

M A Bennett, L S Petro, A. T. Morgan, F. De Martino, L Muckli (2016): High-resolution 7T fMRI reveals auditory and imagery information in non-stimulated visual cortex. 21st Ann. Meeting, Org. Human Brain Mapp. June 26-30, Geneva, Switzerland.

M A Bennett, L S Petro, L Muckli (2016): Investigating cortical feedback of background scene and embedded objects to foveal and peripheral V1 using fMRI. Society for Neuroscience, Nov. 12-16, San Diego, USA.

M A Bennett, L S Petro, L Muckli (2015): Investigating scene feedback to foveal and peripheral V1 using fMRI. European Conference on Visual Perception. Aug. 23-27, Liverpool, UK.

M A Bennett, L S Petro, L Muckli (2015): Investigating scene feedback to foveal and peripheral V1 using fMRI. Psychology Postgraduate Affairs Group. Jul. 22-24, Glasgow, UK.

L Muckli, L Vizioli, L S Petro, F De Martino, **M A Bennett** (2015): Using high-resolution layer-specific fMRI to investigate predictive coding of auditory and

contextual information in early visual cortex. 45th Ann. Meeting, Society for Neuroscience. Oct 17-21, Illinois, USA.

L Muckli, L Vizioli, L S Petro, **M A Bennett**, F De Martino, P Vetter (2015): Predictive coding of auditory and contextual information in early visual cortex – high-res layer 7T. 21st Ann. Meeting, Org. Human Brain Mapp. June 14-18, Hawaii, USA.

M A Bennett, L S Petro, L Muckli (2015): Investigating scene feedback to foveal and peripheral V1 using fMRI. 21st Ann. Meeting, Org. Human Brain Mapp. June 14-18, Hawaii, USA.

M A Bennett, L S Petro, L Muckli (2015): Investigating scene feedback to foveal and peripheral V1 using fMRI. European Conference on Visual Perception. Aug. 23-27, Liverpool, UK.

M A Bennett, L S Petro, L Muckli (2015): Investigating scene feedback to foveal and peripheral V1 using fMRI. Psychology Postgraduate Affairs Group. Jul. 22-24, Glasgow, UK.

M A Bennett, L S Petro, L Muckli (2014): Investigating the effect of stimulus & task on top-down influences to V1 using fMRI. 20th Ann. Meeting, Org. Human Brain Mapp. June 8-12, Hamburg, Germany.

Acknowledgement

Firstly, I would like to express my gratitude to Prof Lars Muckli for hiring me and for his supervision, invaluable insights, and knowledge. I particularly appreciated the intellectual respect shown to me and the freedom to develop and pursue my own ideas. I would also like to thank Dr Lucy Petro for her ceaseless support and advice throughout the last 4 years.

I also would like to thank my secondary supervisors, Dr Marios Philiastides and Dr Lisa DeBruine, for their continued engagement in my progression, and for encouraging my efforts. Thanks to Frances for saving my data from myself in the early days, and for making scanning a data collection a little less boring.

I thank the ERC for funding, and University of Glasgow, Psychology Department for opportunity to conduct my research at this institution.

Additionally, I would like to thank to my fellow lab colleagues, Fiona, Tyler, Angus, Yulia, Hanna, Michele, Johanna, Grace, Philipp, Gemma and Luca for invaluable discussions, advice, for the memories I have (and don't have) from attending conferences with them and for making work not feel like work.

A huge thanks to Hanna for teaching me more about my own brain as I learned about brains in general …and for not minding when I was in work mode.

Lastly, I would like to thank my family for always encouraging me to continue to learn and for their continued support during these years.

P.S. A special thanks to Leon for hearing my frantic cries for help from inside a locked bathroom on the Sunday night before my corrections were due the next morning. Apologies to the University of Glasgow for breaking the door handle in my bid to escape…

Author's Declaration

I declare that, except where explicit reference is made to the contribution of others, this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

_____

Matthew Bennett

Definitions/Abbreviations

BAC	Backpropagation activated calcium spike firing
BOLD	Blood oxygen level dependent
CBF	Cerebral blood flow
CBV	Cerebral blood volume
EBA	Extrastriate Body Area
ECoG	Electrocorticography
EEG	Electroencephalography
EPI	Echo-planer imaging
FFA	Fusiform face area
fMRI	Functional magnetic resonance imaging
GLM	General Linear Model
hMT	Human motion area
HRF	Haemodynamic response function
LFP	Local field potential
LGN	Lateral Geniculate Nucleus
LOC	Lateral occipital complex
MEG	Magnetoencephalography
MT	Medial temporal area
MUA	Multi-unit activity
MRI	Magnetic resonance imaging
MVPA	Multivariate pattern analysis
PPA	Parahippocampal Place Area
pRF	Population receptive field mapping
RF	Receptive field
ROI	Region of interest
RSC	Retrosplenial cortex

SUA	Single-unit activity
SVM	Support Vector Machine
TOS	Transverse Occipital Sulcus
V1	Primary visual cortex
V2	Early extrastriate visual area
V3	Early extrastriate visual area
VTC	Ventral temporal cortex

1 General Introduction

In this general introduction l will described the organisation of the visual system in terms of its feed-forward and feedback processing streams. I will also describe fMRI and the blood oxygenation level dependent (BOLD) signal and how it may be used to study visual feedback processes. I will conclude with a brief rationale for studying feedback and in particular for why we focussed on feedback to foveal and peripheral early visual cortex.

1.1 The human eye and the feedforward visual system

The task of the visual system is to extract useful information from the pattern of light photons entering the eye. It is fascinating that each individual photon is identical apart from its energy or 'wavelength' and its angle of incidence with respect to the retina but that taken together this forms the basis for our entire visual perception. To understand how this happens in detail is the goal of visual neuroscience.

Photons arriving at the eye from between a 60° angle in the medial direction to a 107° angle in the lateral direction are focused by the eye's lens onto the retina where they stimulate photocells with a vertically flipped image. These photocells come in two varieties: rods and cones. Rods are about 20 times more numerous in the retina and appear at a high density everywhere except for the fovea. Conversely, cones are highly concentrated in the fovea and appear at a low density beyond 10°. Cones are also responsible for our ability to perceive colour because cones are responsive to one of three wavelengths (blue, green or red). The rods and cones send their outputs to ganglion cells (via a network of laterally interconnecting cells) of which there are several different types each with different receptive field sizes and sensitivities to depth, shape and colour. Each photocell responds only to photons arriving from within a small region of the visual field called the receptive field (RF) of the cell. The concept of the receptive field is relevant for almost all of the visual system. Later cells pool inputs from several earlier cells with similar RF locations and thus are themselves selective to a larger, but still localised region of the visual field.

The nerve impulses travel from the retina out of the eye via the optic nerve, with some cross-over at the optic chiasm and onto the lateral geniculate nucleus (LGN) in the thalamus. Nerve fibres originating from the left visual field are projected to the LGN in the right hemisphere and vice versa. The LGN is a six layered structure, with each layer receiving input exclusively from one eye and principally from one of two ganglion cell types (parvocellular and magnocellular). Importantly, within each layer of the LGN, adjacent neurons receive input from adjacent neurons in the retina. That is, there is a topographic mapping between locations on the retina and locations in each layer of the LGN. Many of these features of the LGN are manifested in the cortex.

The LGN relays the signals ipsilaterally to the striate cortex (V1), the earliest cortical stage of visual processing which is located at the posterior tip of the brain and forms the calcarine sulcus and extends up the banks onto the top of the surrounding gyri. About 90% of feedforward retinal input is sent to V1. V1, like most of the cortex, is a six layered structure and the signals from the LGN are projected to cells in layer four. The topographic mapping between the retina and the LGN is retained in the transfer to V1. Therefore V1 of each hemisphere contains a vertically flipped representation of the contralateral visual field with the horizontal meridian represented in the deepest part of the calcarine sulcus. However, owing to the increased density of photocells in the fovea of the retina and the uniform density of the topographically corresponding cells in V1, much of the visual cortex is devoted to processing a small area of the central visual field (this concept is known as cortical magnification). This cortical magnification already suggests a functional dichotomy between processes carried out by foveal and peripheral early visual cortex.

The traditional standard model of 'simple cells' in V1 is given by a linear receptive field (Carandini et al., 2005). This refers to a series of weights to apply to each image pixel before summing to predict a cell's response to the image. The idea of a receptive field implies that all areas outside of a sub-region of the image are set to zero – thus the cell responds only to stimuli within this sub-region. Together the weights define a filter, and the shape of this filter is a description of the selectivity of the cell. That is, an image that resembles the filter will produce a large response. V1 cells are commonly found to have filters elongated along one

spatial axis and are thus sensitive to orientation. In addition, the filters usually have an 'ON region' (in which bright and dark stimuli activate or suppress the cell response, respectively) and an 'OFF region' which has the opposite behaviour. Given these filter preferences, a popular conception of V1 is as a bank of Gabor filters. In other words, V1 starts its processing by detecting low-level features in the image that correspond to edges. Roughly half of V1 cells are well explained by this model but the other half, called 'complex cells', have more complicated filters and display more invariances to stimuli (e.g. bright vs. dark). This marks a trend in the visual system of ever more complex feature selectivity and invariance to stimuli. Complex cell responses to stimuli within their receptive field can even be modulated by the concurrent presentation of stimuli outside their receptive field which would in isolation produce no response (Angelucci & Bressloff, 2006). Angelucci and Bressloff (2006) suggest that feedforward and lateral (i.e. intra-areal) inputs account for the 'classical receptive' field while feedback inputs account for the 'extra-classical' receptive field modulations. To understand feedforward and feedback connections in the cortex, it is necessary to discuss the cortical layers.

Like most of the cortex, V1 is composed of six principal layers. Simple and complex cells are not uniformly distributed across layer. Indeed, the layers are defined based on the different cell type concentrations along with their short and long range connections. To simplify a highly complex system, feedforward input arriving from the LGN enters the cell-dense layer 4 (also called the 'granular layer). Layers 2 and 3 (superficial to layer 4) feedforward signals to later cortical areas, whereas layers 5 and 6 (inferior to layer 4) feedback signals to earlier cortical (and sub-cortical) areas (Markov et al., 2013; K. S. Rockland, 2017). Feedback inputs usually arrive into Layer 1 which is cell sparse but which contains the dendritic tufts of large pyramidal cells in Layer 5 (Larkum, 2013; K. S. Rockland, 2017). There are also feedforward and feedback connections variously originating and terminating in both supra- and infra-granular layers and that these supra- and infra-granular counter-streams are at least as functionally segregated as the feedforward and feedback streams (Markov et al., 2013; Markov et al., 2014). Neurons aligned orthogonally with respect to the layer boundaries form cortical columns. Cells in the same column have matching orientation, spatial

frequency tuning and receptive field locations (although the receptive field sizes tend to be larger outside of layer 4; Fracasso, Petridou, & Dumoulin, 2016).

V1 is the first of many cortical visual areas which together form a hierarchy in which each area performs ever more complex and abstract computations and the RF sizes increase in size (Grill-Spector & Malach, 2004). Mapping from the retina to V1 is topographic in that adjacent regions on the retina project to adjacent cortical regions. Therefore, in the cortex neighbouring portions in the visual field are represented adjacent patches of the grey matter surface. This concept is called retinotopy. Many visual areas inherit the retinotopy of the areas project to it. Most cells in V1 project their output to the adjacent cortical area V2 which in turn principally projects to the next adjacent cortical area V3. Areas V2 and V3 are each formed by a concentric strip of cortex wrapping around V1 at the occipital pole. The lower visual field representation in V1 has the horizontal meridian in the deepest part of the calcarine sulcus and moves to the lower vertical meridian on the border of V1 and dorsal V2 (V2d). This visual field representation is reflected into V2d about its border with V1 and then reflected again into V3d. A corresponding reflection of the upper visual field occurs at the ventral borders of V1, V2 and V3.

Beyond V3, the sequential nature of the processing stream starts to disintegrate, although there are generally two recognised processing streams; the dorsal stream processing spatial location and motion, and the ventral/lateral stream processing identification and recognition (Grill-Spector & Malach, 2004). Located on the ventral/lateral stream are several areas which selectively respond to higher level classes of stimuli (Grill-Spector & Weiner, 2014), such as faces in the fusiform face area (FFA); body parts in the extrastriate body area (EBA); objects/tools in the lateral occipital complex (LOC) and posterior fusiform gyrus; spaces, scenes or buildings in areas like parahippocampal place area (PPA), retrosplenial complex (RSC) and transverse occipital sulcus (TOS).

The traditional hierarchical feedforward processing framework described above is based on much empirical evidence from a wide range of methods and in the past has dominated much theoretical and empirical work. Specifically, stimuli are thought of as undergoing sequential stages of processing - each stage localised to a distinct cortical area and capturing more complex stimuli features than the last

(Felleman & Van Essen, 1991). Ideas based on this framework have been fairly impressive in accounting for cortical responses to simple stimuli, but do less well with more complex stimuli (Carandini et al., 2005).

1.2 Feedback in the visual hierarchy

Local connections (~2mm) dominate the circuitry of the cortex, accounting for more than 70% to a given area, and serve to amplify sparse signals arriving from other areas (Markov et al., 2013). In V1, long-range feedforward thalamic input accounts for only ~5% of excitatory connections with the rest being intracortical, i.e. local and feedback connections (Budd, 1998; Kayser, Körding, & König, 2004). Feedforward and feedback connections constitute distinct neural populations forming two segregated streams with different characteristic connectivity profiles (Markov et al., 2013; Markov et al., 2014). Feedback connections are thought to be somewhat more diffuse than are feedforward (Perkel, Bullier, & Kennedy, 1986; Kathleen S Rockland & Ojima, 2003; Salin & Bullier, 1995) although there is much debate around this issue (Markov et al., 2014). Even if feedback connections are more diffuse than feedforward connections, this is only true on the spatial scale of cell populations (i.e. cortical columns; Salin & Bullier, 1995), and there is nonetheless a reciprocity of feedforward and feedback connections between patches of cortex on a course spatial scale (~10mm). Given the retinotopic organisation of the visual cortex, it follows that feedback connections to V1 originate from populations of cells in higher areas which are sensitive to a larger visual field region centered on the V1 cells receptive fields (Perkel et al., 1986; Salin & Bullier, 1995). The anatomical segregation between feedforward and feedback signals mean that they almost certainly perform different functional roles (Markov et al., 2014). Therefore, studying and understanding feedback in addition to feedforward processing is vital for understanding general brain function.

A functional segregation between feedforward and feedback was shown by Andre Moraes Bastos et al. (2015) who characterised these processes in the cortical hierarchy in terms of the frequency of the neural oscillations. They measured electrocorticography (ECoG, i.e. direct subdural recordings) in monkeys in many visual areas throughout the cortical hierarchy. They then applied Granger causality analysis to determine frequency-specific directed influences between all

pairs of areas recorded. In other words, they estimated the frequency channel that each area used to communicate with all other areas in the hierarchy. Remarkably, they found that feedforward influences (from lower areas of the hierarchy to higher areas) occurred in theta- and gamma- bands whereas feedback influences (from higher to lower areas) occurred in the beta-band. Moreover, the ratio of the feedforward to feedback influences between areas correlated with the ratio of anatomical feed-forward and feedback connections. This result was later replicated in humans (Michalareas et al., 2016) using magnetoencephalography (MEG) - values for anatomical connections were derived from areas in the monkey visual hierarchy for which homologous areas exist in humans. One of the studies included a cognitive task requiring top down control which increased the influence of higher areas over the lower areas. Thus the visual hierarchy involves an interplay of feed-forward and feedback influences which adapt flexibly to meet task demands.

The functional relevance of feedback can also be seen at the cellular level. Angelucci and Bressloff (2006) used single cell electrophysiology to show that feedback connections to V1 are involved in contributing to suppression of V1 cell responses when stimuli are presented to the extra-classical receptive field. Specifically, inactivation of area MT (by cooling) reduced, eliminated, or even reversed the usual suppression effects in V1 cells. Nassi, Lomber, and Born (2013) found the same results when inactivating V2/3 and additionally found that those V1 cells that formally showed the strongest suppression effects showed the largest enhancement with deactivation of V2/3. Lateral connections do not span enough cortical space for adjacent cells synapsing onto the V1 cell to account for these effects, and their slow transmission times rule out the possibility of several cells 'passing on the message'. Conversely, feedback connections easily span the necessary cortical space and have transmission times up to 10 times faster than lateral connections.

Chen et al. (2014) demonstrated that contour integration and background suppression in V1 depend upon recurrent signalling with V4. They conducted single cell studies with more complex stimuli; a large field of randomly oriented line segments forming a "background", with 1, 3, 5, or 7 identically orientated lines positioned in such a way as to form a long (dashed) line running through the field.

V1 and V4 cells with receptive fields centred on the mid-point of the line were simultaneously recorded. They found that V1 first reacted at 44ms, and that V4 reacted at 58ms (14ms feedforward transmission delay). V4 instantly showed contour integration (i.e. increased activity for the 7 lines compared to 1 line occurred at 59ms); whereas contour integration was delayed in V1 until 95ms (i.e. 51ms after V1 began reacting to stimulus onset). Poort et al. (2012) found similar results in monkeys regarding figure-ground segmentation, with the figure boundaries initially detected in V1, and the figure interior enhancement occurring in V4, which was only became apparent in V1 later in time and in infra- and supra-granular layers (Poort, Self, van Vugt, Malkki, & Roelfsema, 2016). In addition, if the figure was task relevant the V1 figure-ground segmentation continued to strengthen with a time-course consistent with feedback. These data corroborate with an initial feedforward sweep of cortical activity starting in V1 and travelling to V4, followed by a second feedback sweep from V4 to V1. Moreover these results were neatly explained by a simple recurrent hierarchical network. Thus it seems that the feedforward and feedback sweeps play distinct functional roles in the ongoing processing of stimuli.

A cellular mechanism that could mediate such a complex interplay of feedforward and feedback influences in an anatomically plausible way was proposed by Larkum (2013). Coincident stimulation of the bodies of pyramidal cells in layer 5 (where feedforward connections are made) and to their dendritic tufts in layer 1 (where feedback is often projected) results in bursts of action potentials (backpropagation activated calcium spike firing or 'BAC') not seen with stimulation to either one alone. Such a 'coincident detection' mechanism between the feedforward and feedback streams allows for amplification (or dis-amplification) of task relevant or predicted feedforward input, or more broadly allows for perceptual context to help disambiguate noisy signals.

A common notion has been that feedforward signals 'drive' activity in subsequent areas in a topographic specific manner and that feedback signals serve as a weaker, top down modulation of the feedforward signals (Andre M Bastos et al., 2012). Such a scheme may account for the effects of large scale activity changes brought about through cognitive acts such as attention to certain stimulus locations or features. However, research is converging (Fang, Boyaci, Kersten, &

Murray, 2008; Murray, Boyaci, & Kersten, 2006) to show that manipulating high level aspects of the stimuli (e.g. surrounding depth cues) can result in differential responses in early visual cortex. According to the traditional hierarchical model, such differential responses are expected to be confined to higher areas. These results implicate the involvement of the higher areas in connection with V1 (i.e. feedback) beyond a simple modulation of a feedforward signal. Recently these results have led to widespread acknowledgment that feedback processing is highly important in explaining cortical responses to complex stimuli, and in explaining response modulation by the surrounding context and by internal processing states, even in "low-level" areas (Muckli & Petro, 2013; L. S. Petro, Vizioli, & Muckli, 2014).

Modern comprehensive theories of cortical function must account for the complex interaction between feedforward and feedback among cortical areas. There are now several, not necessarily mutually exclusive, ideas about general brain function that emphasise the role of feedback signals. Friston (2010) suggests that a unifying principle running through many contemporary theories is the idea that the brain maintains a stable state in the face of sparse, noisy sensory input. This idea is usually formalised as the optimisation of some quantity and Friston (2010) interprets this quantity as being equivalent to the minimisation of discrepancies between ongoing internal states and incoming sensory input. Most theories retain the hierarchical framework. An example is predictive coding (Rao & Ballard, 1999) which posits that the brain continuously generates internal models of the world and the models' predictions constitute feedback signals. These predictions are sent from higher to lower areas in the hierarchy and compared to incoming feedforward signals (which initially were generated by stimuli in the outside world) and discrepancies between the predictions and the feedforward data are computed (so called 'prediction errors'). The prediction errors are sent from lower to higher areas and thus represent the feedforward signal in the cortex. The prediction errors are used to update the models to minimise future error. Predictive coding theory has recently received detailed theoretical treatment in relation to current anatomical knowledge (Andre M Bastos et al., 2012) with the conclusion that the circuitry of the cortex appears well-suited to implementing some form of predictive coding. These theories are abstract descriptions of complex and messy real-world phenomena (i.e. the functioning brain) and it is

important to attempt to bridge the gap. In addition to anatomical considerations this endeavour will require measurement techniques and paradigms that are capable of disentangling feedback from feedforward influences.

1.3 Using fMRI to study feedback

Functional magnetic resonance imaging (fMRI) can non-invasively measure the oxygenation of blood (this is known as the blood oxygenation level dependent or 'BOLD' signal), which itself is an indirect measure of neural activity. The BOLD effect comes about through a still not clearly understood interaction of haemodynamic changes in cerebral blood flow (CBF), cerebral blood volume (CBV) and metabolic changes within neural cells. The most relevant unknown is how various neural activities (different cell types, excitatory vs. inhibitory, feedforward vs. lateral vs. feedback connection types etc.) couples and interacts with vasculature effects, since this limits our understanding of which aspects of the BOLD signal are most informative about neural activity. Nonetheless, the value of being able to non-invasively measure some aspect of neural activity in the awake human brain is clear, especially given that relatively large volumes (up to the entire volume of the brain) can be recorded in one or two seconds. This allows a signal time series to be acquired from many brain areas simultaneously. The signal is spatially localised to a few cubic millimetres or less (depending on the field strength of the magnet) which allows an analysis of individual brain areas and in many cases of sub regions of areas. This gives fMRI a unique advantage – no other in vivo imaging method gives the kind of global coverage at such high spatial resolution. However, the signal is temporally sluggish compared to the speed at which much of cortical dynamics unfolds - the BOLD signal peaks around 4-6 seconds after a burst of neural activity. This necessitates various experimental design considerations (for instance the rapidity at which stimuli are presented and so forth).

A key validation of the idea that the BOLD signal is coupled to neural activity comes from studies in which both BOLD and direct local measures of electrophysiological activity are recorded in the same subjects in the same locations of the brain. Jacques et al. (2016) did just this using ECoG measurement of the ventral temporal cortex (VTC) in awake human subjects. A typical BOLD activity profile in response to high level visual stimulus categories (e.g. faces,

body parts, objects and buildings) was observed and a similar profile was also observed in the ECoG data and displayed a clear spatial correspondence. Moreover, the representational structure of distributed VTC activity was similar in the BOLD and ECoG data. Such a correspondence between BOLD and ECoG has also been observed by other researchers in different areas (including motor cortex, sensorimotor cortex and motion area hMT) using various stimuli (Gaglianese et al., 2017; Hermes et al., 2012; Siero et al., 2013; Siero et al., 2014). In general the finding across these studies is that the ECoG activity >60 Hz can be used to best predict the BOLD signal.

The relationship between cerebral hemodynamic changes and the underlying neural activation is complex. There are many aspects of neural activation and the extent to which each of these aspects is represented by the BOLD signal is important for the functional interpretation of the results of fMRI experiments. A key question that has received considerable treatment is the contribution of neural input (dendro-soma activity), and output (action potentials or 'spikes'). Lee et al. (2010) used optogenetics to induce spiking in a population of pyramidal cells in the motor cortex of adult rats while simultaneously recording the BOLD signal. They found positive BOLD activity in the stimulation site and in the region where the pyramidal cells projected to. However, Logothetis (2010) points out that pyramidal cells are known to drive local microcircuits and thus the initial spiking would immediately initiate a cascade of dendro-soma activity. Logothetis and colleagues have tackled this issue by simultaneously recording BOLD fMRI and electrophysiological activity in the visual cortex of awake monkeys undergoing visual stimulation (Goense & Logothetis, 2008; Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001; Magri, Schridde, Murayama, Panzeri, & Logothetis, 2012). The electrophysiological recordings made were: 1) local field potentials (LFP) which reflect an average of neural input (synchronised dendro-somatic postsynaptic activity) a few millimetres in the vicinity of the electrode tip and 2) single and multi-unit activity (SUA and MUA) from individual neurons (mostly larger pyramidal cells) which reflect neuronal output ('spikes'). The general finding was that LFPs (particularly at 20-60 Hz) better predict the BOLD response than does either SUA or MUA. LFPs are even predictive when no spiking activity is present. This implies that the BOLD signal mostly reflects the local neural input to a cortical area rather than the outgoing spikes (although the two will often be correlated).

Therefore the BOLD response may be measured without any outgoing spikes from a cortical area. Magri et al. (2012) used information theory to demonstrate that each frequency band in the LFP contains unique sources of information about the BOLD signal. They speculate that non-feedforward and even non-stimulus-driven components are important in bringing about the BOLD signal.

Harvey et al. (2013) recorded BOLD and ECoG in the same subjects with identical stimuli which were designed to allow an estimation of the receptive fields in V1. The receptive fields estimated from the BOLD data exhibited the familiar centre surround structure, whereby positive BOLD results from center receptive field stimulation and negative BOLD results from stimulation in the receptive field surround. Interestingly, the receptive-field estimated from gamma band activity lacked the suppressive surround, consistent with this frequency's connection with feedforward processing. On the other hand, alpha band activity increased when stimuli were presented in receptive field surround suggesting a role for alpha band activity in producing a negative BOLD response. Given that the BOLD response is not primarily driven by outgoing spikes (feedforward activity), it may be a uniquely sensitive method to study feedback activity in the cortex (Muckli, 2010).

1.4 Visual perception and cortical magnification

One of the most notable features of vision is how a variety of perceptual phenomena vary as a function of eccentricity. A well-studied example (first described in 1923 by Korte) is 'Crowding', which refers to impairment of the perceptual discrimination of a target caused by neighbouring distracter items (Levi, 2008). The critical proximity of neighbouring distracters to the target for crowding to occur varies as a function of eccentricity and can occur with proximities up to approximately half the eccentricity of the target and this rule of thumb may hold all the way to the fovea (Levi, 2008; Strasburger, Harvey, & Rentschler, 1991). In this case, multiple stimuli will be most easily discriminated at the centre of our gaze. Given our cluttered world, this will render most objects unrecognisable to us without fixating them (Levi, 2008).

Such perceptual acuity effects that vary across eccentricities can often be predicted by the number of cortical cells devoted to processing the relevant region of the visual field (i.e. cortical magnification). Moreover, a well replicated finding

is that scaling stimuli to match their cortical representations (known as 'M-scaling') compensates for the usual perceptual degradation of the more eccentric stimulus (Aubert & Foerster, 1857; Cowey & Rolls, 1974; Daniel & Whitteridge, 1961; Drasdo, 1977; Hilz & Cavonius, 1974; Ludvigh, 1941; Rovamo & Virsu, 1979; Rovamo, Virsu, & Näsänen, 1978; Virsu & Rovamo, 1979; Weymouth, 1958).

However, some effects operate differently in the fovea and periphery – for instance, Johnson and Scobey (1980) show that the amount by which a line must be displaced in order to be detected varies as a function of line length beyond 18° (decreasing as line length increases) but is constant when presented foveally. A similar pattern was shown where line luminance as opposed to length was varied. In such examples, M-scaling does not restore behavioural performance. For instance, stimulus contrast interacts with eccentricity when recognising numerical characters (Strasburger et al., 1991; Strasburger, Rentschler, & Harvey Jr, 1994). The key difference may be the need to integrate several features and recognise patterns, rather than merely detect a single feature.

Similar results were obtained by Jüttner and Rentschler (1996) using grating stimuli which varied along two feature dimensions and which were classified by subjects as belonging to one of three sets. In addition to showing once again that M-scaling failed to alleviate the peripheral disadvantage, Jüttner and Rentschler (1996) were able to model the internal stimulus representations of the sets under foveal and peripheral presentations. With this approach, they demonstrated that the behavioural deficit in peripheral presentations was due to these stimuli being represented in a one-dimensional feature space, whereas foveally presented stimuli were represented using a two dimensional space. This is an important finding since it lays bare the extra flexibility of foveal cortex beyond the fact that more neurons are devoted to it. Xing and Heeger (2000) provide a case where M-scaling is completely inconsequential - surround suppression. In surround suppression, the perceived contrast of a central target (in this case a circular sine wave grating patch) is reduced by the presence of a surrounding annulus stimulus (also a since wave grating). They show that this effect is much stronger in the periphery than in the fovea (despite M-scaling) and that discrepancies between the spatial frequency (by 2 octaves) or orientation (orthogonal) of the target and

surround all but eliminated the effect in the fovea but only reduced it in the periphery.

These effects demonstrate a fundamental difference between the fovea and the periphery and highlight the greater sensitivity and precision of foveal cortex. This naturally leads us to ask to what extent these disparate cortical tissues, although both continuously and seamlessly present across most/all of visual cortex and in most individual visual areas, play fundamentally different processing roles.

1.5 Thesis rationale

The rationale for this thesis is to investigate feedback to early visual cortex in human subjects. One of the fundamental properties of the visual system as a whole is its retinotopic organisation. Cortical magnification in early visual cortex means that there is a range of resolution/spatial frequencies processing capabilities which varies smoothly as a function of cortical distance. This suggests that different functional processes are carried out in foveal and peripheral cortex.

It is unclear whether feedback plays different roles in foveal and peripheral cortex. It might be that feedback signals from higher areas regarding computations on data stored at different resolution/spatial frequencies might be directed to sub regions of early visual cortex with the appropriate resolution/spatial frequency processing capability. Evidence for this view would show that early visual cortex plays a more ongoing and active role in perception than simply as an initial low-level pre-processing unit.

We use three main paradigms: 1) Occluding the sub-region of natural scene images in order to isolate feedback signals to corresponding sub-regions of V1. 2) Presenting bistable images in which high and low spatial frequencies give rise to different natural scene percepts. 3) Presenting natural auditory scenes to blindfolded subjects and analysing the visual cortices.

2 Chapter 2: Retinotopic biases in object and scene feedback to V1 are task-dependent

2.1 Abstract

In this chapter I present three experiments which all use a visual occlusion paradigm recently developed in our lab in order to isolate cortical feedback to early visual cortex. The overall aim was to dissociate feedback of object and scene background information to foveal and peripheral regions of V1. To study foveal in addition to peripheral V1 regions, we had to modify the paradigm slightly. The first experiment can be seen as an initial validation of the paradigm modification, using only scene stimuli. The second experiment was our first attempt to include objects as well as scenes using the modified paradigm. The third experiment was an attempt to improve and extend the results from the second experiment. I therefore present the first experiment along with a short discussion of the results and then present the second and third experiments together.

Identifying the objects embedded in natural scenes is a complex problem for the visual system which relies on recurrent processing between lower and higher visual areas. However little is known about how cortical feedback from specialised object and scene areas converges in early visual cortex. Subjects discriminated objects or background scenes in images with occluded central and peripheral subsections, allowing us to isolate cortical feedback activity to foveal and peripheral regions of V1. We found that feedback of object information is projected to foveal V1 cortex. There are also hints that this representation becomes more detailed during an object identification task and that lateral occipital complex could be the source of this increased detail feedback. Background scene information is projected diffusely to all of V1, but can be disrupted by a sufficiently demanding object discrimination task. We suggest that retinotopic biases throughout the visual hierarchy provide an organisational scheme for segregated cortical feedback of information about distinct higher level stimuli.

2.2 Chapter Introduction

One of the classic operations of the visual system is to extract from a single image several well defined pieces of information about its contents. To do this, the visual system is composed of several areas that specialise in the analysis of various categories of stimuli such as faces, objects and places. Somehow, the visual system is able to segment image elements belonging to these categories, which then allows further discrimination of the category instance by the appropriate higher areas. Moreover, different categories can be prioritised for processing according to the task at hand (Cukur, Nishimoto, Huth, & Gallant, 2013).

A particularly well studied process necessary for such specialised image processing to occur is the segmentation of a 'figure' from the 'background'. Recent evidence from neurophysiology and neuroimaging studies suggest that recurrent signalling between higher and lower visual cortex is a key mechanism by which this is achieved. Chen et al. (2014) demonstrated that contour integration and background suppression in V1 depend upon recurrent signalling with V4. Poort et al. (2012) found similar results regarding figure-ground segmentation, with boundaries initially detected in V1, followed by figure enhancement effects fed back to V1 from V4. In addition, if the figure was task relevant the V1 figure-ground segmentation continued to strengthen with a prolonged timecourse consistent with feedback. These data were neatly explained by a simple recurrent hierarchical network. Consistent results were obtained in humans using EEG and similar stimuli by Scholte, Jolij, Fahrenfort, and Lamme (2008). They found that event –related potentials differentiated boundaries first at occipital sites, before progressing to peri-occipital, temporal and parietal sites. In contrast, the figure interior was first differentiated in temporal regions and then propagated back through peri-occipital and finally occipital regions. Thus, recurrent signalling between V1 and higher areas is involved in segmenting objects and scene and this allows specialised processing in higher areas.

V1 and higher areas share several organising principles and these may have consequences for recurrent signalling between areas. One is 'retinotopic organisation', which means that a shift from central to peripheral, or from upper to lower, visual field, corresponds to a shift in the area of visual cortex most

sensitive to that location. This is true of both lower and higher visual cortical areas, but is especially dominant in the early areas. It is interesting to think about what this would imply for the retinotopic specificity of feedback signals. It is the case that feedback connections to V1 originate from populations of cells in higher areas which are sensitive to a larger visual field region centered on the V1 cells receptive fields (Perkel et al., 1986; Salin & Bullier, 1995). Even though feedback connections are more diffuse than feedforward connections on the spatial scale of cell populations (i.e. cortical columns; Salin & Bullier, 1995), there is nonetheless a reciprocity of feedforward and feedback connections between patches of cortex on a coarse spatial scale. Thus on large spatial scales the foveal and peripheral areas of high level cortex should feedback to the foveal and peripheral areas of low level cortex, respectively.

Levy, Hasson, Avidan, Hendler, and Malach (2001) suggested that discrete, localised stimuli may be more suited to processing by foveal cortex whilst continuous stimuli (e.g. natural scenes) are more suited to processing by peripheral cortex. They showed that this organisational scheme can 'overpower' the retinotopic organisation of the higher, but not lower, visual areas. This is intuitive: if one notices an object in one's peripheral visual field, and decides to discriminate its fine details, one would naturally (almost automatically) direct one's gaze towards the object – bringing it into foveal retinotopic cortical space. Thus, feedback of object information might be principally directed to foveal V1 in preparation for likely saccades to the object. This fits in well with current notions of foveal cortex being specialised for local, discrete, high spatial frequency processing (for a review, see Kauffmann, Ramanoël, Guyader, Chauvin, & Peyrin, 2015). In contrast, general scene information is all around us, and does not require the higher foveal resolution to identify its gist (Kauffmann et al., 2015) and so feedback regarding scene gist may reasonably be confined to peripheral V1 cortex. Thus, we hypothesise that there is a fundamental division of labour between feedback signals to foveal-peripheral V1 according to an object-scene principle.

In recent years our lab has developed a paradigm for isolating and studying cortical feedback in the visual system (Muckli et al., 2015; Smith & Muckli, 2010). The paradigm exploits the strong retinotopy of early visual cortex by showing natural scene images with occluded regions. The cortical representations of these

occluded regions of the visual field thus receive no informative feedforward stimulation from the images used in the experiment. If information about the image can nonetheless be extracted from the cortical representation of the occluded region, we know that this information could not have arrived into this region by the feedforward pathway. It follows that some non-feedforward pathway must be at play - i.e. lateral or feedback connections. By taking care to restrict the ROI such that it represents the innermost part of the occluded region, we decrease the plausibility of lateral intra-V1 connections as an explanation as such connections do not span the necessary cortical distance to link the feedforward stimulated to the occluded ROI. Using this paradigm, we have previously shown that cortical feedback constructs representations of stimuli in a peripheral region of primary visual cortex that is not stimulated by feedforward input (Muckli et al., 2015; Smith & Muckli, 2010). A similar analytical approach has been used to show that feedback transmits information about objects presented to the periphery to non-stimulated foveal but not peripheral cortex (Williams et al. 2008). Follow-up studies provided evidence that this was important for perception using TMS (Chambers, Allen, Maizey, & Williams, 2013) and visual masking (Fan, Wang, Shao, Kersten, & He, 2016).

2.3 Experiment 1

2.3.1 Introduction

We wanted to study feedback to foveal cortex in addition to peripheral cortex using the occluder paradigm. However, in previous experiments we effectively removed occluded foveal voxels from consideration in our analyses. This was due the retinotopic mapping stimuli (see figure 2-1 B left most images) and subsequent thresholding procedures used to extract voxels representing the occluded lower right quadrant, which limited the ROI to the more peripheral portion – away from the borders of the occluder and the image.

We therefore first wanted to extend the occluder to cover both foveal and peripheral visual field. This experiment was designed to test the effect of this manipulation. Therefore, we varied which part(s) of the image are occluded: 1) only the lower right quadrant; 2) only central; or 3) both the lower right quadrant

and central regions. This allows us to answer some important preliminary questions. For instance we can ask what the effect of occluding the entire central visual field in addition to the lower right quadrant is on scene feedback to the periphery. It could be that preventing feedforward processing in foveal cortex (a proportionally large cortical area due to the cortical magnification factor and capable of high resolution processing) affects the scene representation in higher areas, and so critically affects or disrupts feedback signals to early visual cortices. The stimuli in this experiment additionally allows us to test whether the feedback pattern observed in a given ROI when only the corresponding image region is occluded (e.g. the peripheral ROI with only the lower right image quadrant occluded) is generalisable to the pattern observed in the same ROI when the extra occluder is present (e.g. the peripheral ROI with the lower right quadrant AND the central image region occluded).

Assuming the extension of the occluder does not disrupt the feedback, an additional aim is to determine whether scene feedback can be detected in the fovea. Given the peripheral bias of higher order scene areas (Hasson, Harel, Levy, & Malach, 2003; Hasson, Levy, Behrmann, Hendler, & Malach, 2002; Levy et al., 2001; Levy, Hasson, Harel, & Malach, 2004; Malach, Levy, & Hasson, 2002) we hypothesised that feedback of scene information would only be projected to peripheral V1 and therefore only in this ROI would we see scene feedback.

2.3.2 Methods

2.3.2.1   Subjects

Nine healthy subjects (5 male; 18-36 years old) were recruited using the University of Glasgow, School of Psychology subject pool. All subjects had normal or corrected eye-sight and no history of brain damage. Each subject completed two consent forms to ensure that they met the safety criteria for fMRI scanning and understood of the experimental conditions. One subject was excluded because the SVM classification of feedforward data consistently failed, suggestive of large and frequent eye movements.

2.3.2.2 Stimuli and Apparatus

The stimuli were projected, at a refresh rate of 60 Hz and a resolution of 1024 x 768, onto a screen (19.0° × 14.2° visual angle) which subjects viewed via a mirror attached to the head coil. The experiment was programmed and displayed using Presentation (Version 16.5, Neurobehavioural Systems). There were nine experimental conditions and four retinotopic mapping conditions. In addition, subjects underwent polar angle and eccentricity retinotopic mapping to localise V1.

Subjects were shown one of three greyscale static natural scenes ('Road', 'Mountain' or 'Beach') with the central region (3.8° x 3.8°), the lower right quadrant (9.5° x 7.1°), or both regions occluded by a light grey patch. Thus there were nine conditions of interest for later analysis (see figure 2-1 A). Within each occluder configuration, the SHINE matlab toolbox (Willenbockel et al., 2010) was used to match the luminance histograms across the non-occluded regions of the three scenes (i.e. scale the pixel intensity such that there are same number of pixels at each intensity for each scene – this also matches the average/global luminance).
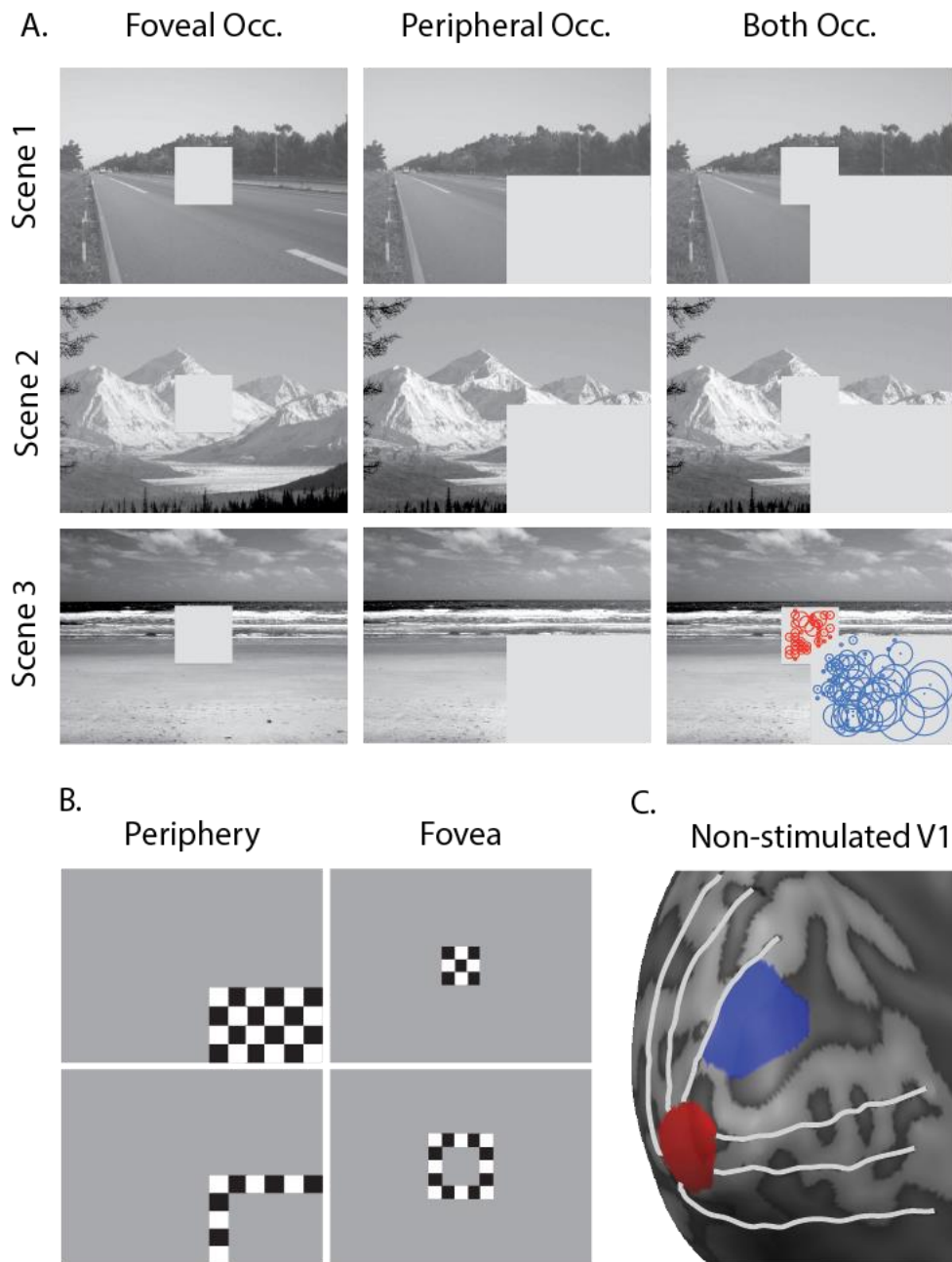
Figure 2-1 A. The experimental images, with an example of the estimated receptive fields of the ROIs overlaid. **B.** ROI mapping stimuli. For each occluded ROI, we contrasted the checkerboard in the top with the checkerboard in the bottom row to produce an initial ROI as shown in **C.** We then further restricted this ROI to those voxels with pRFs falling entirely within the occluded region.

2.3.2.3   Task and Procedure

Subjects completed 5 functional runs. As a task to ensure attention, subjects indicated whether the scene was slightly blurred (8 blurred frames per run, subjects used the index finger of the right hand). Each run consisted of 27 condition blocks followed by 8 retinotopic mapping blocks. Each block lasted for 12s. The condition order was pseudo-randomised for each subject with the

constraint that no condition could be repeated back-to-back. Each block consisted of the scene being flashed on and off (4Hz) for 1s followed by a 1s fixation period. A 12s baseline period during which subjects maintained central fixation preceded each block. Importantly, the occluder itself did not flash on and off, and was present throughout the block. In addition both the central and peripheral occluders were also present during the baseline periods. Therefore, voxels with receptive fields on the occluder during a block were not stimulated in a feedforward manner before, during or after a given block.

The retinotopic mapping blocks (central, central-surround, periphery, periphery-surround) each consisted of 12s of contrast reversing checkerboard stimuli (4 Hz) (see figure 2-1 B). A 12s baseline period, during which subjects maintained central fixation, preceded each checkerboard stimulation block (and also immediately followed the last block). Each run lasted 14min 12s. In addition to the runs described above, polar angle and eccentricity retinotopic mapping were acquired. The scanning session lasted approximately 1.5 hours.

2.3.2.4 Data Acquisition

Functional and anatomical MRI data was acquired using a 3 Tesla MRI system (Siemens Tim Trio) with a 32-channel head coil. For the functional scans an echo-planar imaging (EPI) sequence was used with the following parameters: 18 slices, aligned with the calcarine sulcus, gap thickness 0.3mm, TR-1s, TE-30ms, 426 volumes per run (polar angle mapping = 808 volumes, eccentricity mapping = 552 volumes), a FOV of 220mm, flip angle of 62° and a resolution of 3.0mm^3. The anatomical MRI sequence had a TR of 2.3s, 192 volumes, and a resolution of 1.0mm^3.

2.3.2.5 Data preprocessing

The functional and anatomical data were preprocessed using BrainVoyager QX 2.8 (Brain Innovation, Maastricht, The Netherlands). The first two volumes of each functional run were discarded to avoid saturation effects. The functional data for each run were corrected for slice acquisition time and head movements. Linear and low frequency drifts in the data were removed. The functional data were then aligned with the high resolution anatomical data and transformed into Talairach

space. A model cortical surface of the white matter grey matter boundary was created from the Talairach anatomical scans.

2.3.2.6   ROI Definitions

For each subject, the functional data were projected onto the cortical surface. After localising V1, using the polar and eccentricity data, the ROIs were defined by contrasting the appropriate target mapping stimuli with the corresponding surround mapping stimuli (see figure 2-1 B). Custom Matlab algorithms were used to estimate the population receptive fields (pRF, see Dumoulin & Wandell, 2008) from the polar angle and eccentricity data for all V1 and early visual foveal voxels, allowing us to reject from the occluded ROIs voxels with pRFs falling outside the occluded regions (see figure 2-1 A, Lower right image).

2.3.2.7   Data Analysis

For each Subject, a GLM was used to estimate each voxel's hemodynamic response function (HRF) amplitude for each block. To classify a scene, an SVM classifier was used create a discriminating function for the instances of the three scenes presented, based on the associated multivariate voxel response patterns from 4 runs ("Training"). The discriminating function was then used to classify the scene instances presented during the remaining run ("Testing"). This procedure was repeated, each time using a different run as the "test run" (i.e. 'leave one run out' cross-validation procedure). We also conducted 'cross-classification' analyses in which the conditions on which the SMV was tested and trained were different. Successful classification in this case demonstrates a certain amount of generalisability in the patterns elicited by the conditions used for train and testing.

Since the sensitivity of the SVM varies with the number of features entered, it was important to precisely control for the different number of voxels in the foveal and peripheral ROIs (108.9±21.3 std.dev, peripheral: 48.7±52.7 std.dev, respectively). For each ROI we therefore took equally sized random samples of voxels and ran the SVM. We did this 1000 times and averaged the SVM results to get a single accuracy value. The number of voxels sampled was set per subject as 75% of the number in the peripheral ROI (since this was always the smallest ROI). This

sampling approach has the added advantage that our SVM accuracy estimates are less susceptible to being skewed by any particular combination of voxels.

To assess the significance of group level effects, the mean group classification value was bootstrapped (with replacement) 10'000 times and if 95% of these values exceed chance level (50%), then the classification was considered successful.

2.3.3 Results

As shown in figure 2-2, the scene could be classified above chance from occluded early visual foveal voxels and also from occluded peripheral V1 voxels when a single occluder was present (figure 2-2, comparison 1). Thus, scene feedback is projected to foveal and peripheral cortex. Next, we attempted these same classifications when both occluders were present. The results were the same as before, showing that feedback to peripheral V1 cortex is not critically dependant on feedforward processing occurring in early foveal cortex (figure 2-2, comparison 2). To determine whether the feedback patterns were generalisable with and without the additional occluder (thus preventing feedforward foveal and peripheral processing) we trained and tested the SVM with only a single occluder and tested and trained on the condition with both occluders present. The classification was still successful, demonstrating that the pattern was generalisable (figure 2-2, comparisons 3 and 4). The only indication that the extension of the occluded region altered the feedback pattern was that we could also classify above chance the presence of the second occluder (figure 2-2, comparison 5) – however the global luminance also differed between these conditions and may have driven the classifier result in this case.
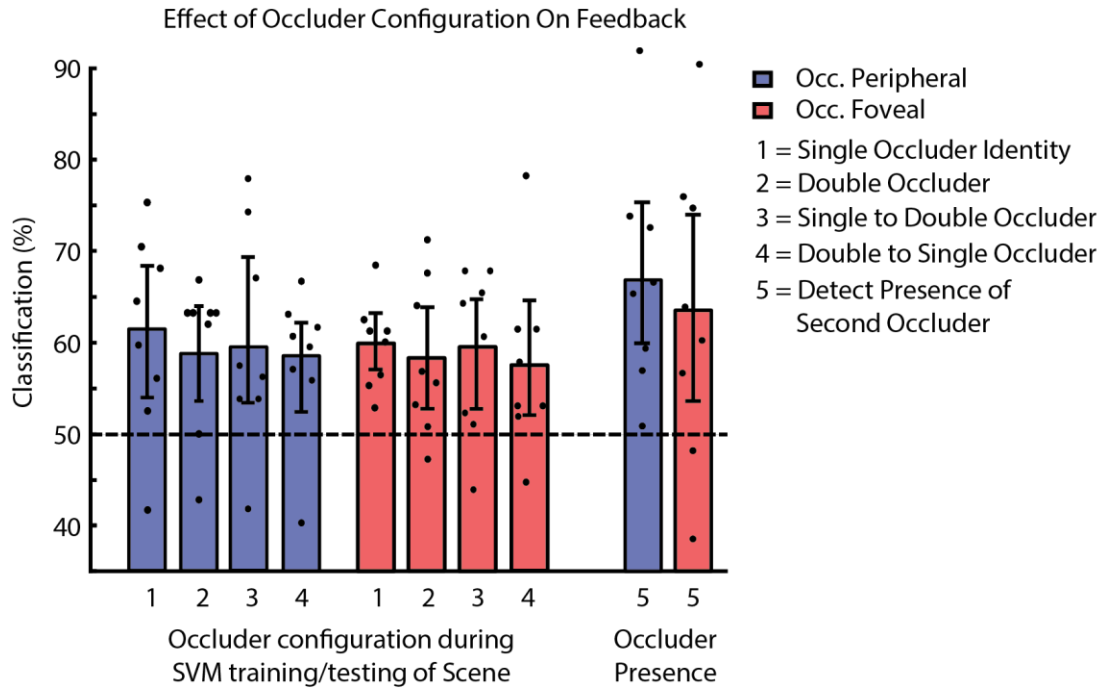
Figure 2-2 Classification in non-stimulated foveal and peripheral V1 cortex in experiment 1. Bars labelled 1 and 2 are regular classification of the scene with single or double occluders. Bars labelled 3 and 4 show cross-classification between single and double occluders. Finally, bars labelled 5 are classification of the presence/absence of the additional occluder. Black circles indicate individual subjects. Errors bars represent 95% confidence intervals obtained from bootstrapping the mean group classification.

2.3.4 Discussion

We now know that both central and peripheral regions can be simultaneously occluded without disrupting the previously established scene feedback effects. We also found, for the first time using the occluder paradigm, that scene feedback is also projected to foveal cortex. This latter result was not expected, based on the peripheral bias of higher scene areas (Hasson et al., 2003; Levy et al., 2004). General discussion of this result is postponed until after experiments 2 and 3 in which we seek to replicate this effect.

The important result from this experiment is that we are now in a position to proceed with an additional central occluder and are thus able to study feedback to foveal and peripheral V1. In subsequent experiments we use stimuli with dissociable object and scene components. We hope to show that object information is fed-back exclusively to foveal cortex, and also to replicate our

previous finding that background scene information is fed-back indiscriminately to all of early visual cortex.

2.4 Experiments 2 and 3

2.4.1 Introduction

Cortical feedback of information about objects embedded in scenes has never been tested. However retinotopic eccentricity biases provide an organisational scheme for segregated feedback of distinct higher level stimuli to early visual cortex. Both higher order and early visual cortex are unified by a large-scale, anatomically predictable, retinotopic eccentricity organisation (Hasson et al., 2003). Moreover, object regions (medial fusiform gyrus, lateral and inferior occipital sulcus) exhibit a central visual field bias, whereas scene regions (PPA and transverse occipital sulcus) exhibit a peripheral visual field bias (Hasson et al., 2003). Thus if object and scene areas send feedback to early visual cortex with similar retinotopic eccentricity biases (as do other visual areas, Salin & Bullier, 1995), then foveal and peripheral V1 cortex ought to receive feedback information about object and scene respectively.

This idea is consistent with existing data. Using fMRI and multivariate pattern analysis (MVPA), object feedback can be detected in non-stimulated foveal cortex, but not in non-stimulated peripheral cortex (Williams et al., 2008) and is related to subject's object discrimination performance (Chambers et al., 2013; Williams et al., 2008). Conversely, natural scene information can be detected in non-stimulated peripheral cortex (Muckli et al., 2015; Smith & Muckli, 2010). We already know from the first experiment that scene feedback is projected to both foveal and peripheral cortex which argues against a strict dissociation of object and scene feedback based on eccentricity. However, this may not be true if an object is also present in the image. Since object feedback is biased towards the fovea, scene feedback may become restricted to the periphery in this case. It is also possible that the effect is modulated by whether the object or the scene is task relevant. It is therefore unknown if there is a general bias for object and scene feedback to be projected to early foveal and peripheral cortex, respectively. If such a bias could be demonstrated, it would imply the existence

of a feedback system organised according to the retinotopy of functionally specialised areas. Also unknown is whether any bias in feedback processing is sensitive to the task-relevance of different image components. Such a task dependant bias would suggest that early visual cortex plays a proactive and functionally flexible role in the disambiguation of image components, beyond the classical view of passively pre-processing and relaying sensory input to higher areas.

We tested whether object and scene feedback are directed to foveal versus peripheral V1 and how task influences this processing. In two experiments, we presented images containing objects, background scenes, or a combination of the two (a similar approach to Harel, Kravitz, & Baker, 2013). By occluding central and peripheral subsections of the image, we isolated feedback activity in the corresponding foveal and peripheral subsections of V1. In the first experiment, similar objects as used by (Williams et al., 2008) were superimposed onto unrelated background scenes. In the other experiment, objects appeared embedded in congruent naturalistic scenes. In both experiments, subjects performed an object or scene discrimination task.

2.4.2 Methods

Unless otherwise stated, the details of experiment 2 and 3 were identical to experiment 1.

2.4.2.1   Subjects (Experiment 2)

Ten healthy subjects (4 male; mean 22.9 years old, range = 18-29).

2.4.2.2   Stimuli and Apparatus (Experiment 2)

There were four experimental conditions and five retinotopic mapping conditions. In addition, subjects underwent polar angle and eccentricity retinotopic mapping to localise V1. The images were one of two grayscale natural scenes ('Mountain' and 'Seaweed') with a pair of superimposed abstract objects belonging to one of two object categories ('Cubic' and 'Smooth', figure 2-3). Preliminary behavioual testing indicated that presenting the objects atop black circles improved

perceptual discrimination performance. In addition, this made our object stimuli more similar to (Williams et al., 2008) in which objects were presented on a uniform black background. Similar to (Williams et al., 2008), each object appeared at 6.1° eccentricity. The width/height of the objects was approximately 2.0° and the black disk was 3.5° in diameter.
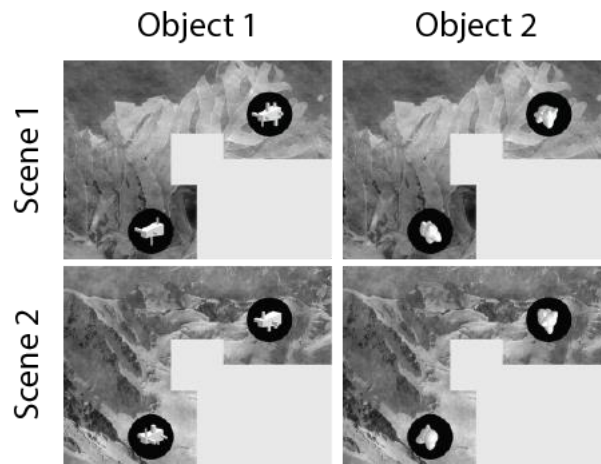
Figure 2-3 Examples of the experimental images used in experiment 2.

The objects were created using a custom algorithm written in Matlab. Specifically, each object shown throughout the experiment was unique (i.e. no object appeared more than once: 1728 unique objects were created in total) and generated by manipulating 4 main feature dimensions (mainly related to the number, size, shape and rotational position of the objects' protrusions). Thus, objects could not be discriminated by looking at only a small part of each object, and more than one location of the object had to be taken into account to attain good discrimination performance. We aimed to make them similar to those used in (Williams et al., 2008). Each image spanned 19° x 14.2°. We occluded central (3.8° x 3.8°) and lower right (9.5° x 7.1°) image portions. To control for low level image properties in those areas of the scenes visible to subjects, the SHINE matlab toolbox (Willenbockel et al., 2010) was used to match the luminance histograms (i.e. scale the pixel intensity such that there are same number of pixels at each intensity for each scene – this also matches the average/global luminance) and the amplitude spectrum of each image was replaced with the average spectrum of the two images, thus exactly matching the images at each spatial frequency and orientation.

2.4.2.3   Task and Procedure (Experiment 2)

Subjects completed 12 functional runs across two days (6 runs per day). In odd/even runs (counterbalanced across subjects), subjects performed one of two tasks: 1) judging whether the two objects were identical or not (50% identical, subjects used the index and middle finger of the right hand, mapping was counterbalanced across subjects) or 2) indicating whether the scene was slightly blurred (8 blurred frames per run, subjects used the index finger of the right hand). Subjects found the object task quite demanding, scoring 76.6% (± 13.2% stderr) correct. In the scene task, subjects detected the blurred frames only 74.4% (± 7.8% stderr) of the time, with a false alarm rate of 25.1% (± 3.8% stderr). However, subjects easily identified the scene: 92.1% (± 0.8% stderr) of responses following true and false blur detections were correct.

Each run consisted of 12 blocks. Each block consisted of 8 trials drawn from the same condition (e.g. Mountain + Cubic objects). The block condition order was pseudo-randomised for each subject, with the constraint that no condition was repeated twice in a row. Each trial consisted of a stimulus flashed on and off (4 Hz) for 1s followed by a 1s fixation period during which subjects responded, thus each block lasted 16s. A 12s baseline period during which subjects maintained central fixation preceded each stimulation block. At the end of the last stimulation block, the retinotopic mapping period started. The retinotopic mapping period consisted of 10 blocks, each lasting 12s, of contrast reversing checkerboard stimuli (4 Hz) at one of five locations (central, central-surround, periphery, periphery-surround, object locations). A 12s baseline period during which subjects maintained central fixation preceded each checkerboard stimulation block (and also immediately followed the last block). Each run lasted 9mins 55s. In addition to the task runs described above, each scanning session ended with either polar angle or eccentricity retinotopic mapping. Each scanning session lasted approximately 1.5 hours.

2.4.2.4   Data Acquisition (Experiment 2)

Experimental runs were 588 volumes per run (polar angle mapping = 808 volumes, eccentricity mapping = 552 volumes) and a resolution of $3.0mm^3$.

2.4.2.5   Data preprocessing (Experiment 2)

The functional data were aligned with the high resolution anatomical data and transformed into ACPC space. A model cortical surface of the white matter gray matter boundary was created from the ACPC anatomical scans.

2.4.2.6   Data Analysis (Experiment 2)

For each Subject, a GLM was used to estimate each voxel's HRF amplitude for each block. We attempted to classify either the scene or object identity using a linear Support Vector Machine (SVM) classifier. This analysis was performed independently for the scene and objects task runs. Specifically, we trained an SVM to classify the Scene or Object using the associated multivariate voxel response patterns. For example, to classify the Scene we labelled the 4 conditions according to the background scene so as to create a binary classification problem (i.e. Mountain + Cubic, Mountain + Smooth labelled '1' vs. Seaweed + Cubic, Seaweed + Smooth labelled '2'). A corresponding approach was taken when classifying Object Identity. We used a 'leave one run out' cross-validation procedure. The number of voxels in the ROIs were foveal: 98.3±41.1 std.dev std.dev and peripheral: 49.0±16.7 std.dev. These disparities were dealt with as in experiment 1.

2.4.2.7   Subjects (Experiment 3)

Nine healthy subjects (1 male; mean 22.9 years old, range = 20-26) took part.

2.4.2.8   Stimuli and Apparatus (Experiment 3)

The stimuli were projected at a resolution of 768 x 768 onto a screen (19.9° × 19.9° visual angle). There were eight experimental conditions and four retinotopic mapping conditions. Our images (see figure 2-4) were created by combining one of three background possibilities (Beach, Forest, None) with one of three object possibilities (BBQ, Tent, None). Thus there were nine possible image combinations. This resulted in eight experimental images and one blank image used as a fixation condition (None/None). Each image was circular, with a radius of 9.9°. All nine images were occluded in a similar way to experiment 1 and 2

(central radius 2.6° and upper right quadrant). The mean luminance of every image was the same (85). In particular, in each image the mean luminance of the pixels belonging to the background scene and object were also the same (85).
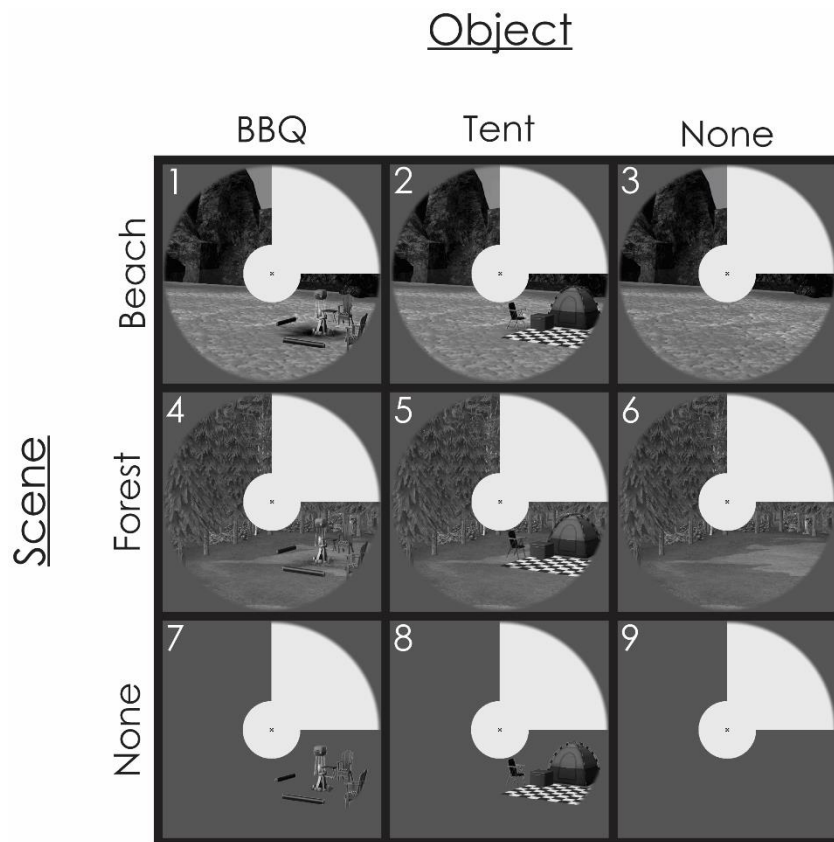
Figure 2-4 The eight experimental images used in experiment 3.

2.4.2.9   Task and Procedure (Experiment 3)

Subjects completed 10 functional runs across two days (5 runs per day). Subjects performed either an object or scene identification task on alternate days (counterbalanced across subjects). The object task was to identify the object in the image (BBQ, Tent, or None present). Similarly, in the scene task subjects had to identify the background scene in the image (Beach, Forest or None present). Subjects made button presses with the index, middle or third finger of the right hand (mapping counterbalanced across runs each day for each subject). Subjects responded at any point during the 12 second block. In contrast to experiment 2, in experiment 3 subjects found both tasks easy: in the scene task scoring 97.5% correct (± 1.6% stderr) and in the object task scoring 96.4% (± 2.1% stderr).

Each run consisted of 24 blocks. Blocks consisted of a stimulus flashed on and off (4 Hz). Subjects responded any time after the onset of a block. The occluder did not flash, and remained present throughout the entire run. The retinotopic mapping period consisted of 4 blocks, each mapping one of four locations (central, central-surround, periphery, and periphery-surround). Each run lasted 11mins 24s.

2.4.2.10   Data Acquisition (Experiment 3)

For the functional scans an echo-planar imaging sequence was used with the following parameters: 31 slices, aligned with the calcarine sulcus, gap thickness 0.2mm, TR-2s, TE-30ms, 342 volumes per run (polar angle mapping = 396 volumes, eccentricity mapping = 268 volumes), and a resolution of 2.0mm3.

2.4.2.11   Data preprocessing (Experiment 3)

The functional data were aligned with the high resolution anatomical data and transformed into ACPC space. A model cortical surface of the white matter gray matter boundary was created from the ACPC anatomical scans. For averaging and performing group statistics on the representational similarity surface maps, we used BrainVoyager QX 2.8 (Brain Innovation) to perform cortical based alignment of all subjects' model cortical surfaces of the white matter gray matter boundary. This creates a mapping from each subject to a common cortical space, based on the pattern of cortical folding.

2.4.2.12   ROI Definitions (Experiment 3)

Since we did not have reverse-checkerboard mapping stimuli for the object locations, V1-object ROIs were defined by contrasting object only experimental conditions with all the other conditions.

2.4.2.13   Data Analysis (Experiment 3)

A Support Vector Machine (SVM) was used to classify the Scene, Object Presence or Object Identity in the VI ROIs. For example, to classify the Scene we labelled the 6 conditions that contained a background scene so as to create a binary classification problem (i.e. Beach, Beach + BBQ, Beach + Tent labelled '1' vs.

Forest, Forest + BBQ, Forest + Tent labelled '2'). A corresponding approach was taken when classifying Object Identity. In the case of classifying Object Presence, we labelled the two object-free scenes as '1' and then classified against the same two scenes with an embedded object. We did this independently for the presence of BBQ and Tent in the scene. The number of voxels in the ROIs were V1-Object: 228.0±27.2 std.dev, foveal: 475.2±117.0 std.dev and peripheral: 116.6±41.1 std.dev. These disparities were dealt with as for experiment 1.

We assessed the similarity in representational geometry between the V1 ROIs and higher areas across the brain (see figure 2-5). To do this, we used a novel analytical approach which combined multivariate searchlights (Kriegeskorte, Goebel, & Bandettini, 2006) and the cross-correlation of ROI representational dissimilarity structures (Kriegeskorte, Mur, & Bandettini, 2008).
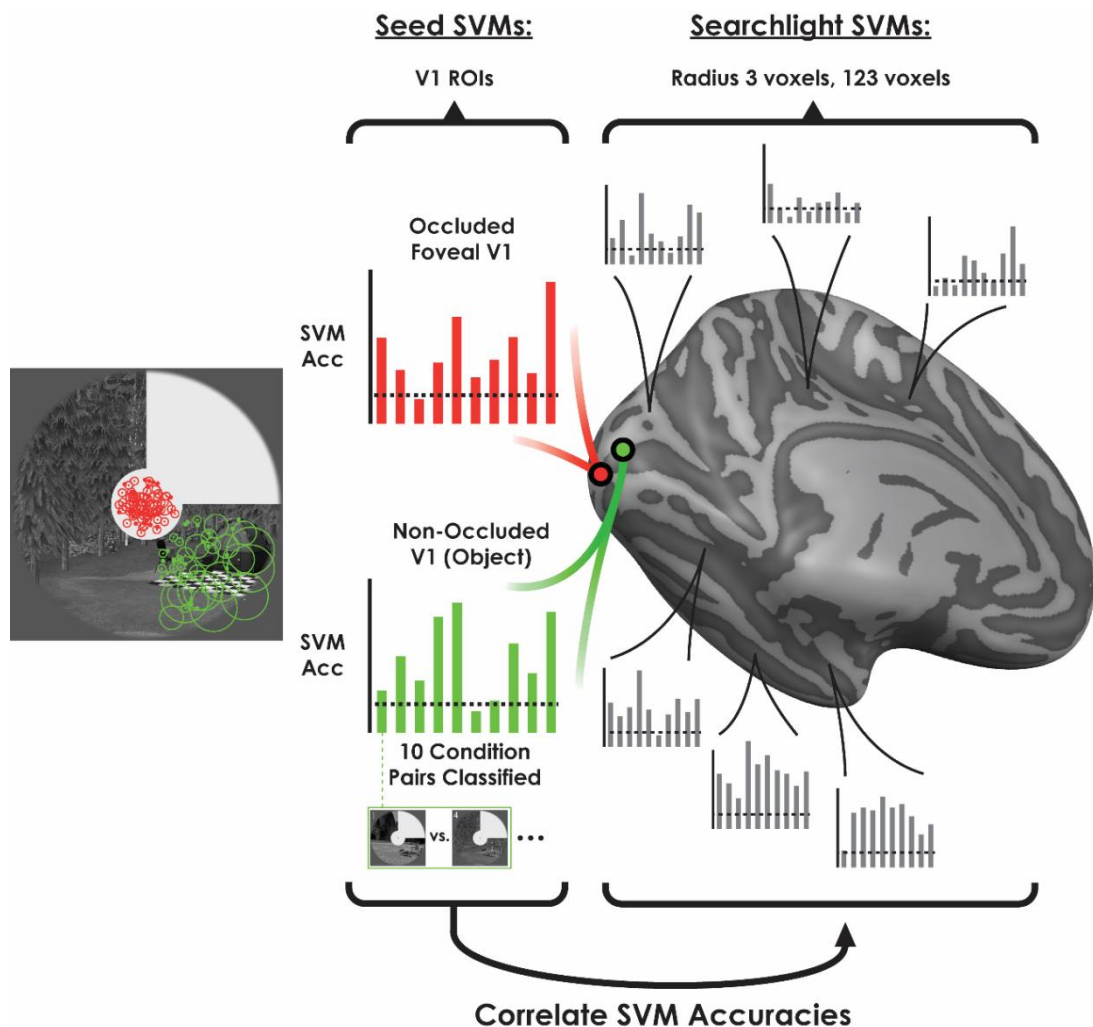
Figure 2-5 Ten pairwise SVM classification accuracies from the V1 ROIs V1-Object, occluded foveal and occluded peripheral (not shown here) were correlated with the ten corresponding SVM classification accuracies obtained from whole brain searchlight maps at each surface vertex.

For each subject, for each task, we ran ten pairwise SVMs variously classifying scene identity, object identity or object presence (in figure 2-4, condition pairs classified were: 1-2, 1-3, 1-4, 2-3, 2-5, 3-6, 4-5, 4-6, 5-6, 7-8) in a searchlight manner. While other SMVs were logically possible to construct, we only chose SVMs in which either the object OR the scene differed between image pairs, but not both, in keeping with our previous analyses in this chapter (hence any difference in classification accuracy can be interpreted as purely object or scene information). The searchlights were performed in volume space for every voxel in gray matter (defined as any functional voxel with at least 1mm3 falling within 3mm of the white-gray matter boundary, on the gray matter side of the boundary). Each searchlight was spherical and centered on a functional voxel with a radius of three functional voxels resulting in 123 voxels per searchlight (no random voxel sampling was used here; unlike for the V1 seed ROIs). The central voxel in a searchlight is assigned the SVM classification accuracy. Thus ten SVM classification maps were created for the entire cortical sheet. The resulting maps were projected onto a cortical surface (for each volume map, the average SVM accuracy occurring 0-2mm along a vertex normal was mapped to the surface vertex). At each surface vertex, we constructed a vector from the ten SVM classification accuracies. This vector characterises the representational geometry regarding the experimental stimuli. We then correlated (using spearman's rho) the vectors of each vertex with a corresponding vector obtained from a V1 'seed' ROI (V1-Object, occluded foveal, occluded peripheral, the random voxel sampling approach was used here as described above). The result was a wholebrain correlation map showing the similarity in representational geometry to the V1 seed ROI. We carried out the above procedure separately for data from the scene and object task runs to assess whether the similarities in representational geometry between the V1 seed ROIs and the rest of the brain were modulated by task.

We used cortical-based alignment to transform subjects' maps into a common space and display them on a group average mesh. This allowed us to average the maps across subjects and to perform group statistics. To generate p-value maps, for each ROI and task, the average representational similarity maps were bootstrapped (with replacement) and the p-value at each surface vertex was calculated as the proportion of bootstrap samples below or equal to zero. Group

difference maps for each ROI were constructed by subtracting the average object task map from the average scene task map. To statistically assess these group difference maps we used a non-parametric permutation approach: We pooled the nine subjects' scene and object task maps (thus creating a pool of 18 maps), and randomly sampled (with replacement) two sets of nine maps, averaged each set, and then took the difference of these average maps. We repeated this procedure 1000 times to build a null distribution of difference maps against which we compared the observed difference map. A *p*-value map (uncorrected) was calculated at each surface vertex as the proportion of null distribution samples with an absolute value below or equal to the observed difference absolute value.

2.4.3 Results

2.4.3.1 Occluded V1 (Experiment 2)

We studied feedback of object and scene information to V1 by analysing multivoxel information patterns in the cortical representations of occluded subsections of foveal and peripheral V1. This analysis was performed independently for the scene and objects tasks.

Subjects found the object task quite demanding, scoring 76.6% (± 13.2% stderr) correct. In the scene task, subjects detected the blurred frames only 74.4% (± 7.8% stderr) of the time, with a false alarm rate of 25.1% (± 3.8% stderr). However, subjects easily identified the scene: 92.1% (± 0.8% stderr) of responses following true and false blur detections were correct.

We were able to decode scene information during the scene task but not during the object task in both foveal V1 (Scene Task: 54.6%, p=0.0022; Object Task: 49.6%, p=0.6547) and peripheral V1 (Scene Task: 54.8% p<0.0001; Object Task: 52.3%, p=0.1387, figure 2-6). In the case of foveal V1, we found that correct classification rates were significantly higher during the scene task (p=0.01), as determined by a permutation test. This result suggests that scene information is fed back to foveal and peripheral V1, in agreement with experiment 1, but that the demanding task of perceptually discriminating the objects disrupts the representation of scene information.
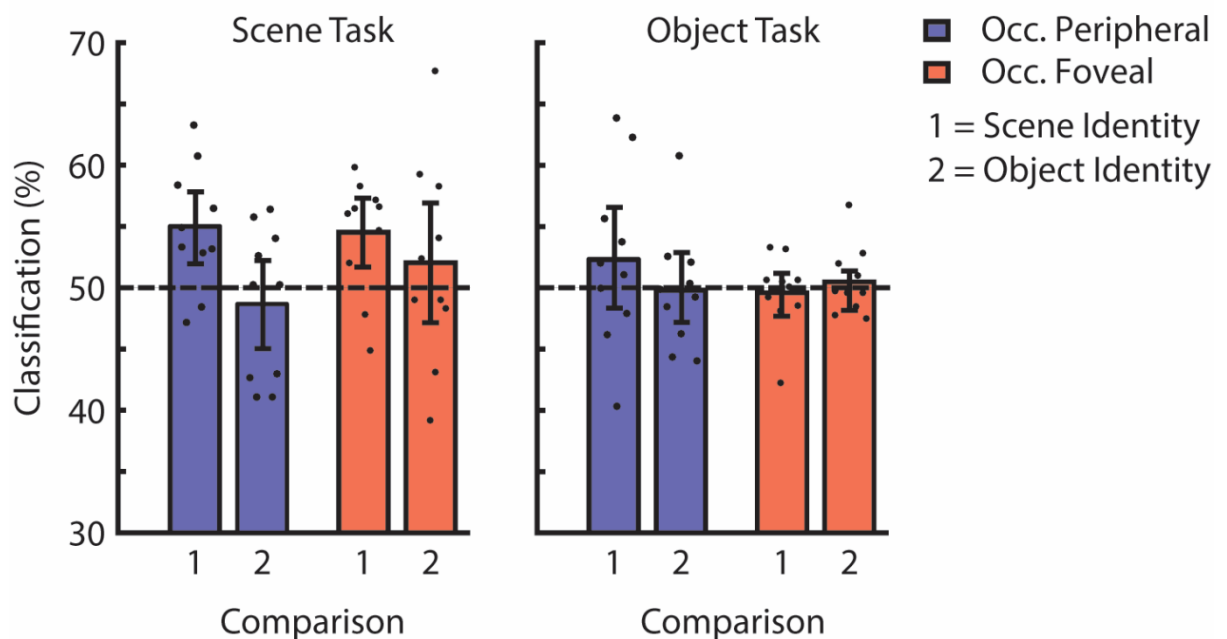
Figure 2-6 Classification in non-stimulated V1 cortex in experiment 2. Black circles indicate individual subjects. Errors bars represent 95% confidence intervals obtained from bootstrapping the mean group classification.

We were not able to detect object identity information in either task in foveal (Object Task: 50.6%, p=0.2410; Scene Task: 52.2%, p=0. 1938) or peripheral V1 (Object Task: 50%, p=0.548; Scene Task: 48.7%, p=0.7501, figure 2-6). This is in contrast to Williams et al. (2008) who were able to decode object identity information in the fovea. In our data, even classification in a V1 ROI directly stimulated in a feedforward manner by the objects was low and only significantly above chance during the object task (Object Task: 52.1%, p=0.0027; Scene Task: 48.6%, p=0.9571). Given that the objects were unnatural in appearance and superimposed onto the scene in a superficial way, and given that discriminating the objects disrupted scene feedback, we reasoned that the scene feedback may

have obscured object feedback. To investigate this possibility, we conducted experiment 3.

2.4.3.2 Occluded V1 (Experiment 3)

In experiment 3, we used computer generated grayscale images of real objects embedded in scenes in a natural way, for example, a barbeque on a beach. With the extended stimulus set, we were additionally able to classify the presence of an object in the scene. Since no fine details are required to indicate an object's presence/absence in the image, this "object presence" feedback information may be easier to detect than the identity – perhaps even so in the periphery. These analyses were performed independently for both the scene and object task data.

In contrast to experiment 2, in experiment 3 subjects found both tasks easy: in the scene task scoring 97.5% correct ($\pm$ 1.6% stderr) and in the object task scoring 96.4% ($\pm$ 2.1% stderr). Scene information was detectable regardless of task in both foveal (Object Task: 55.0%, $p<0.0001$; Scene Task: 55.0%, $p<0.0001$) and peripheral V1 (Object Task: 55.3%, $p<0.0001$; Scene Task: 56.5% $p=0.0017$, figure 2-7). Thus, reducing the object task load (as compared to experiment 2 in which subjects only scored 76.6% correct) may have enabled automatic scene feedback to all of V1. This once again replicates the main findings of experiment 1.
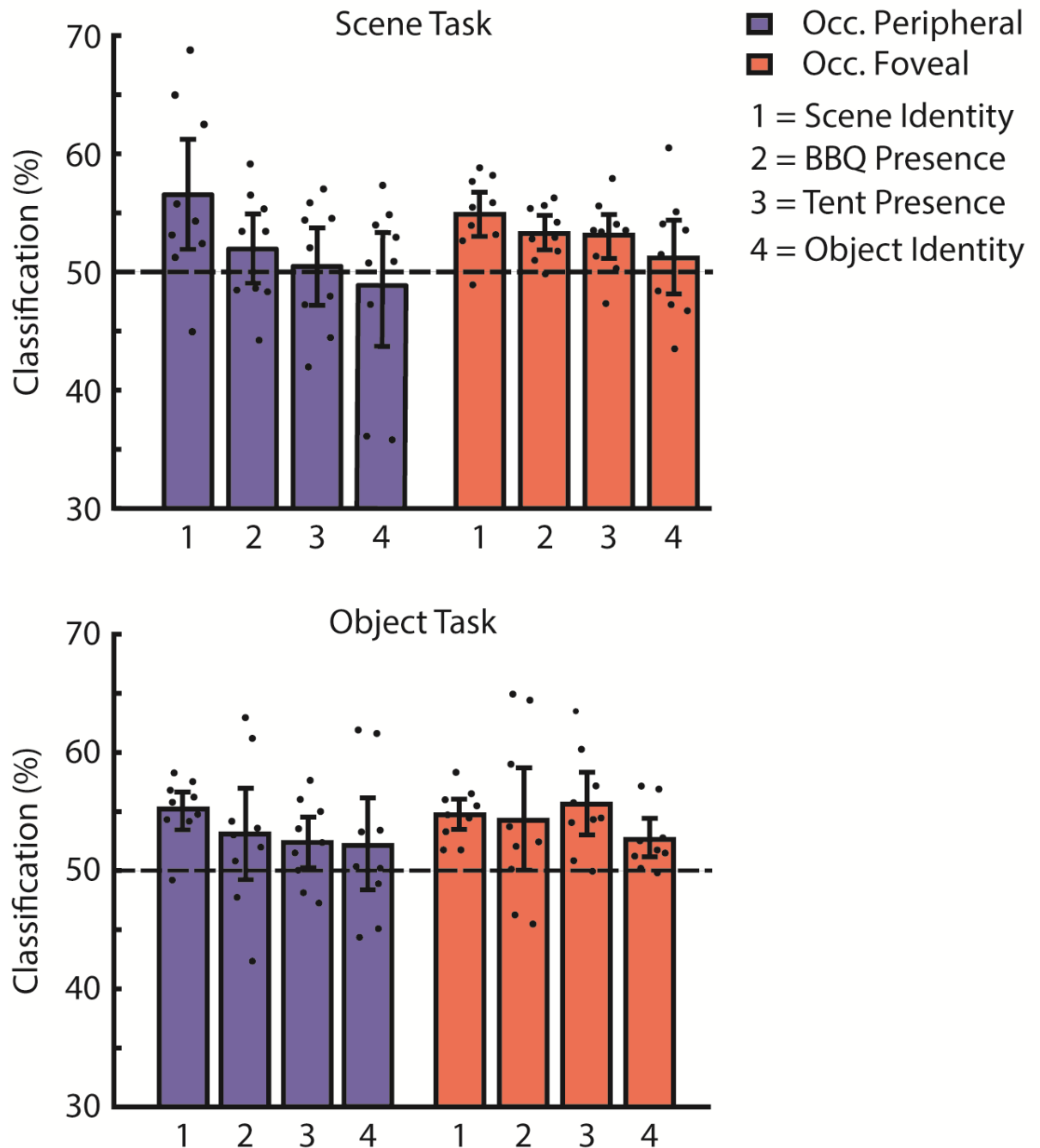
Figure 2-7 Classification in non-stimulated cortex in experiment 3. Black dots indicate individual subjects. Errors bars represent 95% confidence intervals obtained from bootstrapping the mean group classification value.

Object presence in the scene was detectable for both objects, regardless of task, in foveal V1 (BBQ during Object Task: 54.1%, p=0.03; BBQ during Scene Task: 53.2%, p<0.0001; Tent during Object Task: 55.5%, p<0.0001; Tent during Scene Task: 52.9%, p=0.0014, figure 2-7). In peripheral V1, during the scene task, no object information was detectable (BBQ during Scene Task: 51.9%, p=0.1113; Tent during Scene Task: 50.6%, p=0.3592). The presence of objects in the scene was

detectable during the object task for one object, although only at an uncorrected alpha level, and approached significance for the other (BBQ during Object Task: 53.1%, p=0.0572; Tent during Object Task: 52.4%, p=0.0147). However, this weak effect was not supported by further analyses; running the SVM on data pooled across task (which allows the SVM to take advantage of more data) yielded non-significant results for the presence of both objects (BBQ: 50.7%, p=0.2978; Tent: 48.8%, p=0.8829), as did separate SVMs classifying object presence/absence within each scene rather than collapsing across the scene (BBQ: 51.7%, p=0.2363; Tent: 51.8%, p= 0.2012) and also when using Linear Discriminant Contrast (BBQ: 0.2332, p=0.7099; Tent: 0.9703, p=0.1662, Walther et al., 2016). In general there were no significant differences between foveal and peripheral ROIs for object classification accuracies (p>0.05, as determined by a permutation test). However, we do see a similar trend as Williams et al. (2008) which suggests that information about the presence of an object in the scene is automatically fed back to foveal V1.

Object identity information was detectable in foveal V1, only during the object task (Object Task: 52.3%, p<0.0001; Scene Task: 51.1%, p=0.2546). Such information was not detectable in peripheral V1 in either task (Object Task: 52.1%, p=0.1416; Scene Task: 48.9%, p=0.6530, figure 2-7). This suggests that object identity is fed-back to fovea when subjects are asked to identify the objects in the scene.

There were no significant differences in classification rates between ROIs or between different tasks (as determined by permutation tests). This could be due to the low overall classification accuracies across conditions – particularly since subject variation in *below chance* classifications are not interpretable and serve only to add noise to the data. Nonetheless, motivated by a weak but consistent trend for the object classifications to be higher in foveal V1 and (an even weaker trend) to be higher during the object task, we conducted follow-up exploratory analyses to look for differences between ROIs and between tasks.

2.4.3.3 Relation of V1 to Higher Areas (Experiment 3)

Given that the object and scene information we detected in our occluded V1 ROIs likely originates from higher areas, we assessed the similarity in representational

geometry between the V1 ROIs and the rest of the brain, including other visual areas higher up in the cortical hierarchy. We were particularly interested in whether these geometry similarities were modulated by task, given that object identity was only decodable during the object task in foveal V1. To do this, we used a novel analytical approach (figure 2-5) similar to Henriksson, Khaligh-Razavi, Kay, and Kriegeskorte (2015) which combined multivariate searchlights (Kriegeskorte et al., 2006) and the cross-correlation of ROI representational dissimilarity structures (Kriegeskorte et al., 2008).
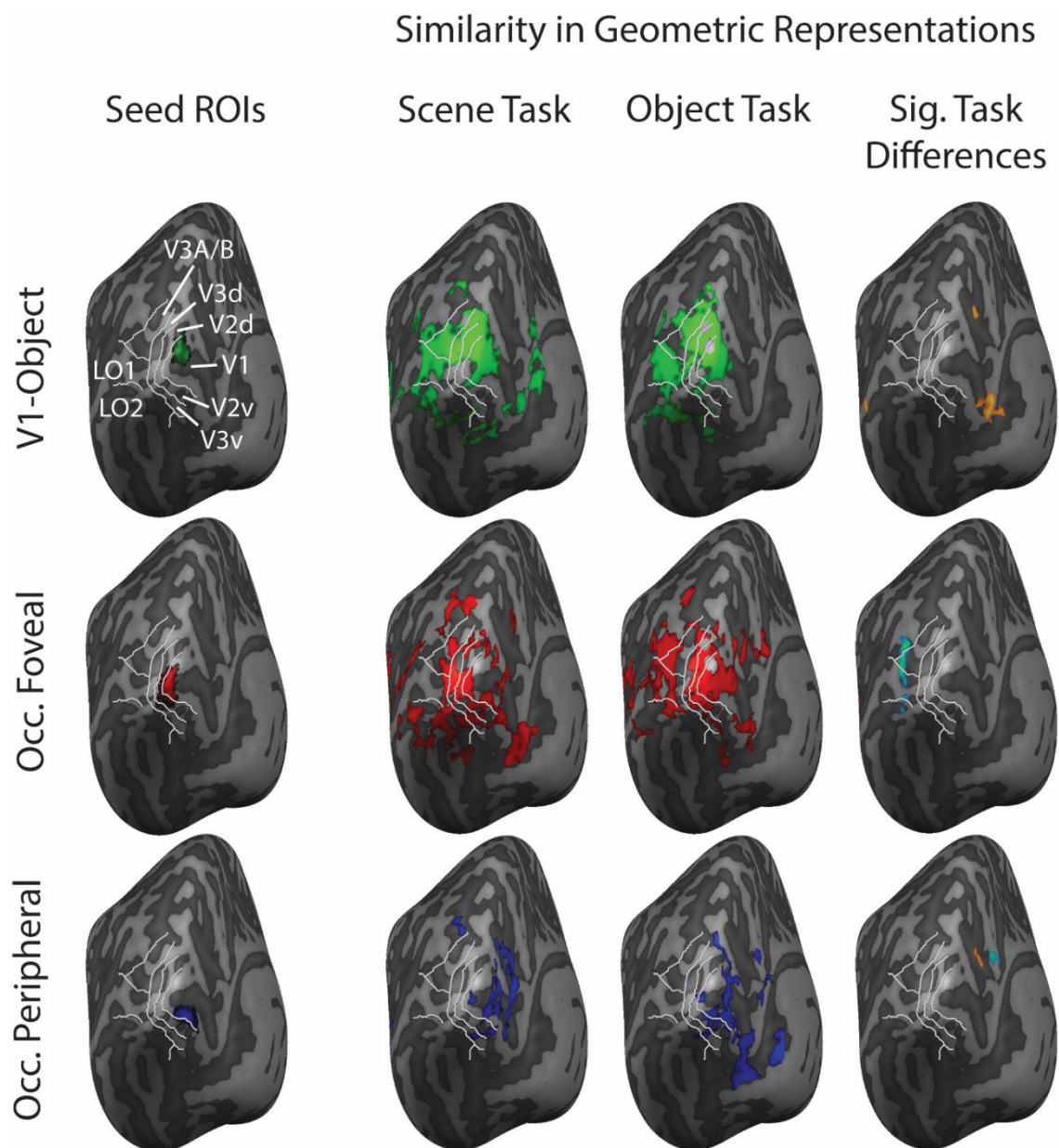
Figure 2-8 The 1st column shows the location of the seed ROIs. The 2nd and 3rd columns show significance values from the whole-brain correlation maps. These maps indicate significant similarity in representational geometry to V1 seed ROIs during the scene and object tasks. The 4th column shows significant task differences between maps (yellow-orange hues and green-blue hues indicate significantly higher correlations during the scene task and object task, respectively).

First, we looked at the feedforward V1-Object ROI (figure 2-8, top row). In both tasks, similar geometries are found in dorsal V2 and V3. This is sensible, because there are strong feedforward connections from dorsal V1 to these regions (Van Essen, Newsome, Maunsell, & Bixby, 1986). Also in both tasks, Lateral Occipital Complex (LOC) showed a similar geometry. Again, this is sensible given LOC's role in object recognition (Grill-Spector, 2003; Grill-Spector, Kourtzi, & Kanwisher, 2001) and the feedforward input of object information to the seed ROI. However, only during the scene task were there additional similar geometries spanning anterior of the parieto-occipital sulcus to parahippocampal place area (PPA). This likely reflects task dependant engagement of scene selective cortex. Non-parametric permutation tests showed significantly higher correlations (uncorrected) during the scene task in RSC and in an area slightly dorsal to parahippocampal place area (PPA).

Next we looked at occluded foveal V1 (figure 2-8, middle row). As this region is occluded, its geometry is based on feedback information, presumably sent from higher visual areas in the cortical hierarchy. In our classification analyses, we saw a trend for more precise object feedback during the object identification task. Consistent with this observation, during the object task the correlation maps show that dorsal early visual cortex and LOC have similar geometries as does occluded foveal V1. Moreover, during the object task this correlation map using occluded foveal V1 as the seed region is strikingly similar to the feedforward V1-Object maps. Consistent with a reduction in detailed object information, the LOC does not appear in the scene task correlation map. Non-parametric permutation tests confirmed that area LOC had significantly higher correlations (uncorrected) during the object task compared to the scene task. However, the effect is small and does not survive False Discovery Rate correction, therefore it can only be taken as a hint towards support for the hypothesis that area LOC, known to be involved in object recognition, may be a source of more precise object information fed-back to occluded foveal V1 when subjects identify objects.

For occluded peripheral V1 (figure 2-8, bottom row), we found significant correlations on the medial surface, but unlike the occluded fovea and V1-object ROIs we found none at all on the lateral surface in either task. This is consistent with the lack of object information shown in our classification analyses in either

task. During the scene task there were significantly higher correlations (uncorrected) in RSC, similar to what we also found using the feedforward V1-Object ROI as a seed region. This may suggest task dependant communication within the cortical hierarchy between scene selective cortex and peripheral early visual cortex. During the object task, there were significantly higher correlations (uncorrected) in a region just anterior to RSC. This latter result is hard to interpret – it could simply be a false positive which serves to remind us to be cautious in drawing strong conclusions on the basis of small effects which may be underpowered.

2.4.4 Discussion

We used occluded images to study cortical feedback of objects and scenes to foveal and peripheral V1. We found that foveal and peripheral V1 contain high level information that could not have arrived directly from retinal and lateral geniculate input. As such, this information must have been projected to V1 from higher levels of the cortical hierarchy, or from lateral interactions within V1. In experiments 2 and 3, we showed that background scene information is fed back to foveal and peripheral V1, replicating the result in experiment 1, and that this can be disrupted by a difficult object task. In experiment 3, we showed that the presence of an object in the scene is fed back to foveal V1 regardless of task. Peripheral V1 on the other hand did not classify above chance in either task. This could be taken to suggest a general bias for object feedback to be projected to early foveal cortex, however without significant statistical comparisons of task this claim can only be a speculation. The identity of the object was only significant in the fovea and then only during the objects task. Again, we are tempted to speculate that the precision of this object feedback to foveal cortex varied with task, but did not find statistically significant differences to confirm this.

We then asked where the feedback ordinates from. We addressed this question using a recently developed analytical approach (Henriksson et al., 2015) which we used to assess the similarity in representational geometry between the V1 ROIs and higher areas across the brain. We found that only during the object task was the representational geometry in occluded foveal cortex similar to LOC. This result supports the idea that LOC may feedback more precise object feedback

information to occluded foveal V1 during the object task. It is important to be cautious in making strong claims about directionality from these results – any similarity in representational geometry between two regions could be mediated by a third region or by an interaction among several other regions. Nonetheless, it is striking that the same areas that are correlated to the feedforward V1-Object seed, regardless of task, are also correlated to the occluded foveal V1 seed during the object task. Task dependant increases in background connectivity between early visual cortex and specialised higher areas (Fusiform Face area and PPA) have also been shown by (Al-Aidroos, Said, & Turk-Browne, 2012).

Given LOC's central visual field bias and known involvement in object recognition (Grill-Spector, 2003; Grill-Spector et al., 2001), feedback of object information to early foveal cortex is anatomically and functionally sensible. The high and low resolution capabilities of early central and peripheral cortex, respectively (Duncan & Boynton, 2003), led Malach and colleagues to propose that the functional relevance of the central-peripheral bias in higher areas may be to accommodate fine detail discrimination of objects and large-scale integration of scenes, respectively (Hasson et al., 2002). However, this account does not address the fact that centrally biased higher areas still have very large receptive fields (Grill-Spector & Malach, 2004), which make high resolution processing difficult. Retinotopically segregated feedback might solve this problem by recruiting V1's highest resolution capabilities (the fovea) to scrutinise and discriminate objects (Hochstein & Ahissar, 2002). Our results are consistent with this proposal, although we find feedback of scene information to foveal as well as to peripheral cortex – we did not find a complete dissociation. It maybe that the quantity and quality of computations that can be performed by foveal cortex is useful for most visual perceptual tasks. Indeed, Larson and Loschky (2009) demonstrated with a behavioural occluder paradigm that once absolute area is equalised, central vision is more efficient than peripheral at recognising scene category. This experiment was modelled using deep convolutional neural networks (Wang & Cottrell, 2016) and a wider stimulus set which included faces and objects as well as scenes with similar results. In addition, Wang and Cottrell (2016) found that faces and objects benefitted more than scene from processing by central visual field and the reverse was true for peripheral visual field. Thus feedback to both foveal and peripheral cortex may aid scene identification. It is possible that more minimalistic scenes

(e.g. simple spatial layouts as used in Harel et al., 2013) would only be fed back to peripheral cortex.

Another reason for object feedback to be restricted to foveal cortex is suggested by neurophysiological monkey data from (Poort et al., 2012) showing that the distribution in visual space of enhanced V1 figure responses predicted saccade land positions. Objects are likely targets for upcoming saccades (Nuthmann & Henderson, 2010), especially if they are task relevant (Henderson, 2003), and thus a possible reason for object feedback to be projected only to foveal cortex may be to meet relevant upcoming feedforward input to V1 resulting from saccades. This hypothesis is not mutually exclusive with that of Malach and colleagues; it may be that when object's identity needs to be scrutinised, foveal V1 cortex is recruited, via feedback, for its high resolution capabilities. This may occur as preparation for a natural tendency for upcoming saccades designed to center the object on foveal V1 cortex. Cortical predictions are fed back to V1 in time for sensory inputs that are the target of eye movements (Edwards, Vetter, McGruer, Petro, & Muckli, 2017), and so predicted object information may be transported to foveal cortex because we tend to make eye movements towards objects to interact with them. The content of the feedback may be sensory predictions (Friston, 2010) of the object representations generated in higher visual areas and sent to early visual areas, such that internal representations of 'to-be-processed' objects are present with the highest resolution in V1.

We highlight the dynamic recurrent information sharing that occurs between V1 and functionally specialised higher areas during the disambiguation and recognition of object and scenes. We propose that the role of V1 should be thought of as a proactive and functionally flexible extension of the higher areas. V1's unique abilities, such as high resolution processing, are recruited adaptively according to task requirements.

2.5 Chapter Discussion

We have shown that feedback signals are present when the visual system analyses natural images. If the image is composed of a background scene and an object, feedback signals related to these image components show a somewhat different

pattern of projection sites. Specifically, object feedback seems directed to foveal cortex while scene feedback is apparently sent to both foveal and peripheral cortex. There were small hints of task effects (mostly not significant) in the classification accuracy results. The comparison of the representational geometries of occluded early visual areas with that of non-occluded category selective higher areas reflected the task effects and allows us to motivate hypotheses about the interplay between areas and how different stimuli and tasks affect those relationships. The idea that retinotopic organisation might play an organisational role in feedback (as well as feedforward) signalling, coupled with the known eccentricity biases in category selective higher areas suggests that object stimuli have some special significance for foveal cortex even in early visual cortex.

Hasson et al. (2003) provide evidence that higher object areas such as LOC are biased to analysing central visual field information whilst higher scene areas such as PPA and RSC have a peripheral visual field bias. Foveal cortex is typically associated with smaller receptive field sizes and thus increased resolution capabilities and the availability of high spatial frequency information. A functional consequence of such a central visual field bias could be that this makes them more suited to analysing small details and thus enable fine discrimination of object features. Conversely, the peripheral visual field is characterised by large receptive fields, low resolution and thus low spatial frequency information (Kauffmann et al., 2015). Hasson et al. (2003) suggest that scene selective higher areas are more likely to integrate information over large regions of the visual field. If (Hasson et al., 2003) are correct, then our object decoding data could be taken to support the hypothesis that higher object areas send feedback to foveal early visual cortex in order to make use of the high spatial frequency information.

It is worth considering that the typical receptive field size of higher order visual areas is several times larger than early visual cortex. Thus any higher areas must rely on early visual cortex to access the higher spatial frequency information. This can be done using feedforward or feedback signals. The fact that foveal visual cortex is large is both a reflection of the fact that higher spatial frequency information contains more detailed information and therefore needs a larger cortical allocation and that the visual system relies heavily on such information for perception. As such it might be that the higher scene areas actively use the

computational ability of early foveal visual cortex well after the initial feedforward signals arrive. Low spatial frequency information may be sufficient for gross scene categorisation, but two scenes will still each contain diagnostic high spatial frequency information.

The feedback processing stream is retinotopically coarser than the feedforward processing stream (Muckli et al., 2015). Therefore, in the case of feedback in the visual system it may be that the high vs. low spatial frequency content of the image is a better driver of foveal vs. peripheral visual cortex than is retinotopic location of the image. In other words, even though there is a correlation between eccentricity and spatial frequency information in visual cortex, these two properties could have a different importance for feedback and feedforward processing. It could be that there is a bias to send feedback about low spatial frequency information to the periphery, and to send high spatial frequency information to foveal cortex – somewhat independently of the eccentricity at which these frequencies are presented. If true, this would mean that the scene feedback we observed in foveal and in peripheral V1 carried distinct information. An interesting follow up experiment to investigate this idea would be to independently filter the background scene and objects. The prediction would be that if the background scene contains only low frequency information, then feedback would be constrained to peripheral V1. If the background scene contains only high frequency information, then feedback would be constrained to foveal V1. In the case of objects, since there is evidence that feedback is sent only to foveal cortex, we would predict that filtering the object of high frequency content would remove the feedback signal entirely. This latter prediction has been demonstrated in a behavioural experiment by (Fan et al., 2016) using the same object stimuli as the original fMRI study by (Williams et al., 2008).

In the next chapter, we sought to test the influence of top down spatial frequency biases in the representation of scene images in early visual cortex.

3 Chapter 3: Classifying category perception of bistable hybrid images in early visual cortex

3.1 Abstract

We present two fMRI experiments in which we present subjects with 'hybrid images', which can be perceived as one of two natural scenes – one carried in low and one carried in high frequency information channels. We find that far peripheral early visual cortex may be influenced in a top down manner, whilst foveal cortex seems to more faithfully represent the feedforward input. This provides evidence for scene related feedback to far peripheral cortex. Both experiments have very few subjects, and so the conclusions are only speculation at this point. However, the theoretical implications are interesting if the observed trends are true.

3.2 Chapter Introduction

In the last chapter, we used a paradigm in which subsections of an image were occluded allowing us to isolate and examine feedback signals arriving into corresponding subsection of V1. This subsection of V1 received no relevant feedforward input. In real-life however, feedback and feedforward signals converge in the same regions of cortex and interact dynamically. Therefore the occluded paradigm is an artificial perceptual situation. In order to study feedback in a more ecologically valid context, it is desirable to have relevant feedforward information. This is challenging since the feedforward signals could mask the feedback signals in the BOLD signal. One way to approach this is to use stimuli with a bistable percept. That is a single stimulus with two mutually exclusive perceptual interpretations. A classic example is the Necker Cube illusion (in figure 3-1).
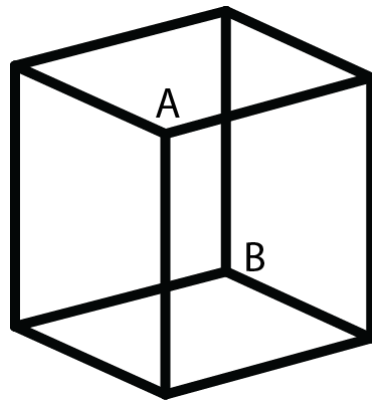
Figure 3-1 The Necker Cube illusion. Vertices A and B perceptually alternate between being perceived as being in the foreground or background, respectively.

The logic is that on any given trial the physical input to the visual system and the corresponding feedforward signal does not explain why the stimulus was perceived in the way it was. Thus any observed differences associated with the different perceptions, even in early visual cortex, can be attributed to top down or feedback influences. Thus we have a way to include the presence of a relevant feedforward signal and still to investigate some aspects of the interacting feedback signal.

Here we use hybrid images which were originally developed by (Schyns & Oliva, 1994). Hybrid images are composed of two superimposed images, the spatial frequencies of each of which is low- or high-pass filtered. This creates a single stimulus image which is bistable – only one of the two filtered images is perceived at any one time. With appropriately short presentation times, subjects tend to perceive just one of the images, with the non-perceived image contributing nothing more than perceptual noise.

In keeping with the experiments in chapter 1, we presented natural scenes from different categories to the entire visual field. Sowden and Schyns (2006) hypothesise that feedback influences can act on spatial information in certain frequency bandwidths and also presented data showing that spatial frequency processing can be modulated in a top-down fashion at early stages of visual processing. Our data from the last chapter suggests that natural scene information is fed-back to both foveal and peripheral early visual cortex, therefore we probed for natural scene feedback at a range of eccentricities. We were interested in

whether feedback attributable to low spatial frequency information would be found more in the periphery and high spatial frequency information would be found in the fovea.

Previous research using hybrid stimuli (Oliva & Schyns, 1997; Schyns & Oliva, 1994; Sowden, Özgen, Schyns, & Daoutis, 2003) were exclusively behavioural and the trials were much more numerous and presented much closer together in time than is usual for fMRI paradigms. We therefore had to modify the behavioural experimental design to be suitable for an event-related fMRI design. Critically, the inter stimulus intervals had to be substantially lengthened and although we observed some behavioural effects, they were not as clear as in previous research. Nonetheless, the results were informative.

Previous studies claim that early visual processes are likely involved in producing the behavioural effects and base this conclusion on the observation that the effects can be targeted to a circumscribed region of the visual field. However, this research only probed such retinotopy at the scale of visual field quadrants. Thus the effects are constrained to any visual area which is selective to the left/right and upper/lower visual field. Many visual areas – several of them high level - fit this constraint. Therefore, using fMRI may help better localise the effect.

3.3 Experiment 1

3.3.1 Introduction

We do not always perceive what is physically in front of us. This has repeatedly been demonstrated using visual illusions, where the experience of physical features are affected and even changed by prior knowledge, context, and expectation (see e.g. Muckli, 2010; Scholte et al., 2008; Wokke, Vandenbroucke, Scholte, & Lamme, 2013). For example, Schyns and Oliva (1994) developed hybrid images, which are two superimposed images of different categories each filtered so as to contribute non-overlapping spatial frequency information (high or low spatial frequencies). These hybrid images give rise to the striking observation that only one image can be perceived at a time, although both images are physically

present (Oliva & Schyns, 1997). By manipulating the presentation time duration, experimenters were able to bias subjects' perception to perceive the image in low spatial frequency (shorter presentation time), or in high spatial frequency (longer presentation time) (Oliva & Schyns, 1997; Oliva, Torralba, & Schyns, 2006; Sowden et al., 2003). Remarkably the experimenters were also able to 'sensitise' the subject (without the subject being aware), to be biased to perceive a particular spatial frequency independent regardless of stimulus duration. This was achieved by presenting a train of 'sensitisation images', which only contained an image in the to-be-sensitised spatial frequency and structured noise (phase scrambled) in the other spatial frequency. Subjects gradually 'tuned in' to the spatial frequency in which meaningful information was present. When hybrid images were occasionally presented, subjects tended to perceive the image in the sensitised spatial frequency without realising that there was also an image in the other spatial frequency. This suggests that top-down expectation plays a role in spatial frequency processing. This leads to the question of how and where these operations are processed in the brain. Of particular interest is whether these 'top-down' influences of expectation are fed back all the way to early visual cortex. To our knowledge, fMRI has not been used to investigate the role of feedback in processing spatial frequencies.

Only a few fMRI studies have investigated spatial frequency processing in the visual cortices and generally used stimuli such as sinusoid gratings or spatial frequency band-passed scene images (Kauffmann et al., 2015). Interestingly, these studies showed that there were differences in how foveal and peripheral visual regions process spatial frequency. Specifically that HSF images tend to activate areas linked to foveal vision, while LSF images tend to activate regions associated with peripheral vision (Musel et al., 2013; Peyrin et al., 2005). This agrees with monkey studies by Tootell, Silverman, Hamilton, Switkes, and De Valois (1988) (using the C-2-deoxy-d-glucose uptake technique) showing that spatial frequency preference in low level (i.e. striate) visual areas varies inversely with eccentricity. In the case of higher areas specialising in scene analysis (such as PPA, RSC, OPS etc.) the picture is more mixed. Peyrin, Baciu, Segebarth, and Marendaz (2004) found that low spatial frequencies activated the right PPA and anterior temporal cortex (also known to be involved in scene analysis). However, Rajimehr, Devaney, Bilenko, Young, and Tootell (2011) showed essentially the opposite result in both humans

and macaques while Zeidman, Mullally, Schwarzkopf, and Maguire (2012) show a high spatial frequency preference for right PPA and a low spatial frequency preference for left PPA. One resolution to this may come from Musel et al. (2014) who showed a succession of filtered images of the same scene going from low-high or from high-low frequencies. PPA, RSC and OPA all activated more for the low-high sequence, suggesting this is more in line with how these areas naturally process scenes (a 'course-to-fine approach'). While the connection between spatial frequency and eccentricity is clear in low level areas, findings concerning higher level areas are more mixed.

It is therefore interesting to speculate on how the high level areas interact with low level areas. These studies all assessed feedforward activity, but not top-down influences. Whether the correlation between eccentricity and spatial frequency biases are constrained to low-level feedforward signals or also affect feedback signals is unknown.

We used the 'bistable' percept of hybrid stimuli to see if the neural representations in early visual cortex corresponded to the perceived spatial frequency (i.e. top-down perceptual experience) or to both (i.e. the physical feedforward activity). In addition, by using sensitisation (Oliva & Schyns, 1997) we aimed to bias the subject's perception to a particular spatial frequency, and to assess if this bias strengthened over time as more sensitisation images were presented. We examined whether these effects were differentially present in foveal and peripheral visual regions, as previous research shows that feedforward processing of spatial frequency bands changes with eccentricity (Musel et al., 2013; Peyrin et al., 2006). We hypothesise that low and high frequency information might be preferentially fed-back to peripheral and foveal cortex, respectively.

3.3.2 Methods

3.3.2.1 Subjects

Five healthy subjects (4 male; mean 26.3 years old, range = 23-31) were recruited using the University of Glasgow, School of Psychology subject pool. All subjects had normal or corrected eye-sight and no history of brain damage. Subjects gave

consent and were screened in accordance with the safety criteria for fMRI scanning. One subject's data was unusable for fMRI analysis due to data corruption.

3.3.2.2 Stimuli and Apparatus

The stimuli were projected with a refresh rate 70 Hz at a resolution of 768 x 768 onto a screen (19.8° × 19.8° visual angle) which subjects viewed via a mirror attached to the head coil. The experiment was programmed and displayed using Presentation (Version 16.5, Neurobehavioural Systems). In addition, subjects underwent polar angle and eccentricity retinotopic mapping to localise V1.

We presented two kinds of stimuli, created in a similar way to Özgen, Sowden, Schyns, and Daoutis (2005). Hybrid stimuli combined two images from different categories, whereas sensitisation stimuli combined a single image with a 'structured noise image'. We used 4 scene categories (mountain, city, forest, and living-room), with 6 exemplars in each. These choices of category was to balance categories that naturally contain much high spatial frequency content (city and living-room) and low spatial frequency content (mountain and forest). We avoided choosing images with large regions devoid of either low or high spatial frequency content so as not to create hybrids with regions containing diagnostic information about only one category.

To create a hybrid image, the Fourier representations of two images from different categories, one low-pass filtered and the other high-pass filtered, were summed (figure 3-2 A & B). In this way, we created a hybrid for every possible between-category image pairing (e.g. a mountain image was combined with all images from the other categories, but not with other mountain images), and for both filtering possibilities (i.e. we created a version in which the first image was low-pass filtered and a corresponding version in which it was high-pass filtered). In total, this resulted in 432 hybrids.

To create a sensitisation image, the Fourier representations of an image and a noise mask, one low-pass filtered and the other high-pass filtered, were summed (see figure 3-2 C & D). The noise masks were created by summing the Fourier representations of two images from different categories after randomly increasing

or decreasing the phase angles by a constant amount (=0.628). We generated 5 sensitisation images per category exemplar per filtering possibility, each time randomly selecting a pair of images to create the noise mask. In total, this resulted in 240 sensitisation images.

A. Hybrid: Mountain in HSF, City in LSF

B. Hybrid: Mountain in LSF, City in HSF
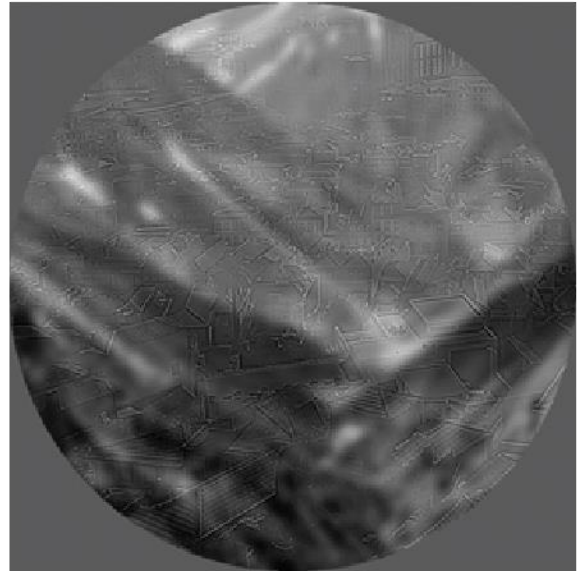
C. Sensitisation: Forest HSF
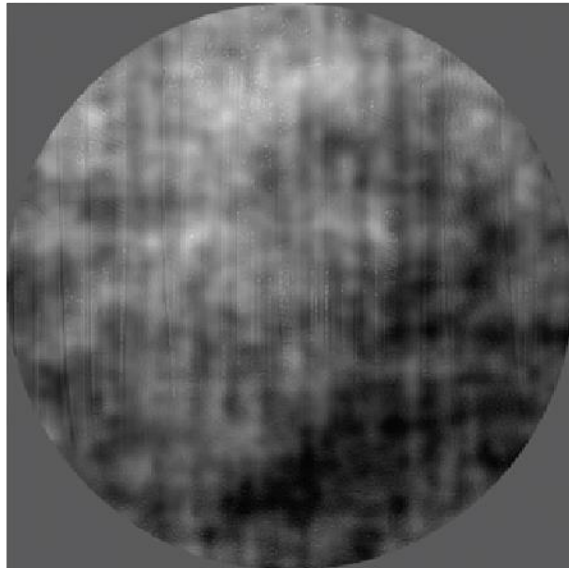
D. Sensitisation: Forest LSF

Figure 3-2 Examples of the experimental stimuli. **A & B. H**ybrid stimuli: A mountain and city scene combined. Panel A shows the mountain scene in high spatial frequency and city scene in low spatial frequency, panel B shows the same images but in opposite spatial frequencies. **C & D. S**ensitisation stimuli: A forest scene in either high spatial frequency (panel C) or low spatial frequency (panel D) combined with a noise image of opposite spatial frequencies.

The distance from the observer to projection screen was 108cm and the diameter of the stimuli was 768 pixels, subtending a visual angle of 19.8°. Hence, one degree of visual angle was 38.8 pixels. The low spatial frequency filter was a 2D symmetric Gaussian with a standard deviation equal to 1 cycle per degree (cpd). The high spatial frequency filter was an inverted 2D symmetric Gaussian with a standard deviation of 5 cpd. The global luminance of the final stimuli were normalised. Moreover, the uniform grey area surrounding the stimuli also had the same global intensity as did the images.

3.3.2.3 Task and Procedure

The experiment consisted of five experimental runs and two retinotopic mapping runs. No hybrid stimulus was repeated during the experiment. The start of each run had a 12 second baseline where only a fixation cross was shown. The fixation cross disappeared when a stimulus was onscreen, so as not to obscure information in the foveal visual field. The subjects were instructed to maintain fixation and indicate which category they had perceived (or which category seemed perceptually stronger if both were perceived – note that this would represent a suboptimal situation from our experimental perspective) with a button press as soon as they could. Each functional run lasted 9 minutes 48 seconds.

Each experimental run began with a block of 24 hybrid stimuli, presented in an event-related design. Each hybrid was shown for 135ms (Özgen et al., 2005) with an inter-stimulus interval (ISI) of 5865ms. The hybrids were pseudo-randomly selected with the constraint that each category was represented twice in each SF and there was an equal number of between-category pairings. After the 24th hybrid image there followed a 10 second baseline.

The second part of the run consisted of 18 sensitisation blocks each followed by a single hybrid. Three blocks sensitising either low spatial frequency of high spatial frequency were presented consecutively. This allowed us to assess strengthening of the biasing as the same spatial frequency was sensitised. The sensitisation blocks consisted of 8 sensitisation images with each category represented twice and shown in a random order. Each sensitisation image was shown for 135ms with an ISI of 1515ms. Following each sensitisation block there was a 5865ms baseline before the hybrid was presented for 135ms. In the 18 hybrid stimuli following the

sensitisation blocks, each category was represented at least once in each spatial frequency. Each run ended with a 10 second baseline period.

3.3.2.4 Data Acquisition

Functional and anatomical MRI data was acquired using a 3 Tesla MRI system (Siemens Tim Trio) with a 32-channel head coil. For the functional scans an EPI sequence was used with the following parameters: 15 slices, aligned with the calcarine sulcus, gap thickness 0.25mm, TR-1s, TE-30ms, 588 volumes per run (polar angle mapping = 792 volumes, eccentricity mapping = 536 volumes), a FOV of 192mm, flip angle of 62° and a resolution of 2x2x2.5mm. The T1 weighted anatomical MRI sequence had a TR of 2.3s, 192 volumes, and a resolution of 1.0mm^3. The T2 weighted anatomical MRI sequence had a TR of 2s, 172 volumes, and a resolution of 1x0.5x0.5mm (later re-sampled to 1mm^3).

3.3.2.5 Data preprocessing

The functional and anatomical data were preprocessed using BrainVoyager QX 2.8 (Brain Innovation, Maastricht, The Netherlands). The first two volumes of each functional run were discarded to avoid saturation effects. The functional data for each run were corrected for slice acquisition time and head movements. Linear and low frequency drifts in the data were removed. The functional data were aligned to the T2 image which itself was aligned to the T1 anatomical, this resulted in a good alignment between the functional and anatomical T1 scans. The run-to-run alignment was assessed using a customised matlab script and thereafter manually evaluated to see if any run was misaligned. None of the subjects' runs were deemed misaligned and thus all runs were included in the analysis. A model cortical surface of the white matter gray matter boundary was created from the ACPC anatomical scans.

3.3.2.6 ROI Definitions

For each subject, the functional data were projected onto the cortical surface. After localising V1, V2 and V3, using the polar and eccentricity data, the ROIs were defined (figure 3-3) by taking eccentricity bands across V1, V2 and V3

approximately corresponding to between 1-3°, 3-6° and 6-10+°. This was done in both hemispheres and the corresponding eccentricity ROIs were combined.

Early Visual Cortex ROIs
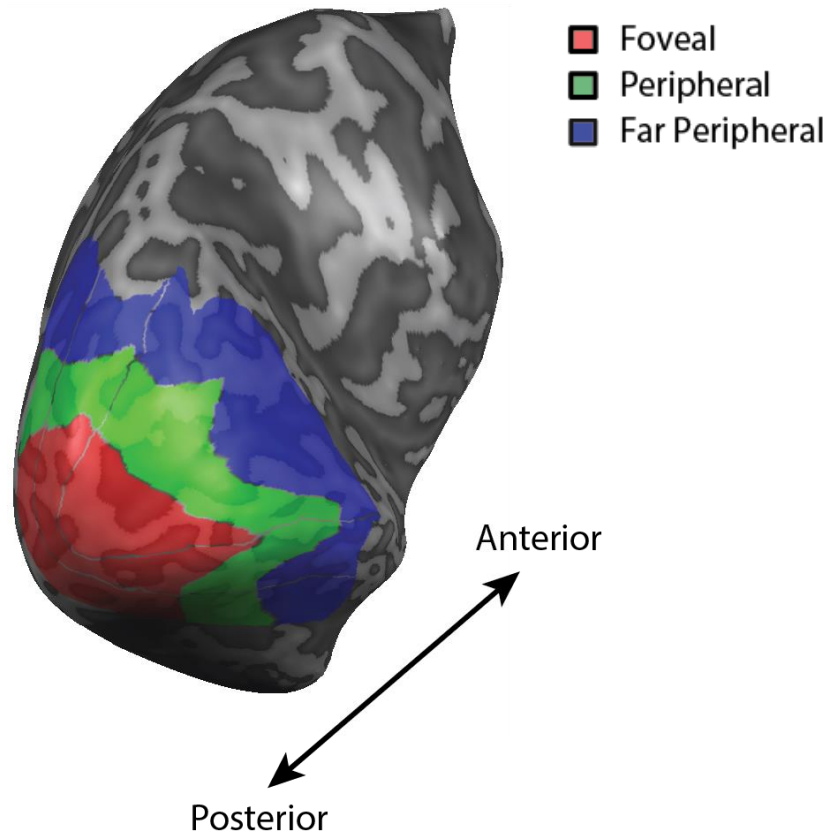
Figure 3-3 Example of ROI definitions in the left hemisphere.

3.3.2.7 Data Analysis

We used a general linear model (GLM) to estimate each voxel's HRF amplitude for each individual event/block in each run. Sixty conditions were defined; the 24 hybrids from the start of the run, each sensitisation block (= 18), and each hybrid following sensitisation blocks (= 18). The behavioural bias modulation due to sensitisation blocks was very weak (i.e. it sensitised trials were not much more likely to be perceived in a given spatial frequency than were non-sensitised hybrids). Thus there were too few sensitised hybrids that followed the expected pattern to support analysis on exclusively the sensitised trials. We opted to pool all hybrid trials (24 + 18) to increase power. To classify trials, an SVM classifier was used create a discriminating function based on the associated multivariate voxel response patterns from 4 runs ("Training"). The discriminating function was then used to classify trials presented during the remaining run ("Testing"). This

procedure was repeated, each time using a different run as the "test run" (i.e. 'leave one run out' cross-validation procedure).

Since a hybrid image consists of two different categories (one in each spatial frequency) we could label all trials according to the category contained in either the low or high spatial frequency. We therefore attempted category classification of the low or high frequency content.

We then attempted to classify all hybrid trials which were labelled according the reported or according to the unreported category, regardless of the spatial frequency of the reported category. We also performed the same analysis for trials in which low spatial frequency or in which high spatial frequency image was reported.

Since the perception was self-reported by the subject in the scanner through button-presses, we also ran a control analysis to evaluate the possible contribution of motor responses to the classification of reported category. To do this we labelled the hybrids according to whether a button press was made with either the right hand (forest and indoors) or left hand (mountain and city). Significant classification for this labelling scheme implies that motor responses may have confounded the results. Additionally, we created labels based on perceptual reports of the man-made (city and indoor) and natural (mountain and forest) categories. In this case, the button presses were more balanced (left 3rd digit + right 2nd digit vs. left 2nd digit + right 3rd digit) and thus ought to be less informative for classification. Therefore, if classification was significant in this instance then the category perception was considered more plausible (albeit indirectly so) than were motor responses in classification of reported category.

To assess the significance of individual subject level effects, we performed a permutation of the classification score for each condition labels (n=1000), and compared this to the real data. The permutation scrambles the labels and yields chance-level results. It was considered significant if the proportion of the permuted scores did not outperform the real scores more than 5%.

To assess the significance of group level effects, we bootstrapped (with replacement) the mean group classification value 10'000 times and if 95% of these values exceed chance level (50%) it was considered the classification successful.

3.3.3 Results

3.3.3.1 Behavioural Results

The behavioural data for sensitisation effect is shown in figure 3-4. Individual data (N=5; the fifth subject still had behavioural data although the imaging data was corrupted) is shown in the first five panels, and the sixth panel shows the group normalised mean data, with bootstrapped (with replacement) 95% confidence intervals. Since each subject was likely to have different inherent perceptual bias towards a spatial frequency, we used all the data from the hybrids only phase at the start of each run to estimate a baseline. Thereafter, in the sensitised hybrid data, we calculated the percentage of responses that the subject reported the spatial frequency to which the block was sensitising towards (sensitisation strength). In the group averages it can be observed that for HSF the perceptual bias has significantly changed from baseline after one and three blocks of sensitisation (block one: 95% CI [0.63 – 25.97]; p=0.01, block three: 95% CI [0.1 – 23.47]; p=0.023). For LSF, the perceptual bias shows a significant change from baseline after sensitisation blocks two and three (block two: 95% CI [-0.8 – 13.96]; p=0.037, block three: 95% CI [-3.06 – 28.2]; p=0.04).
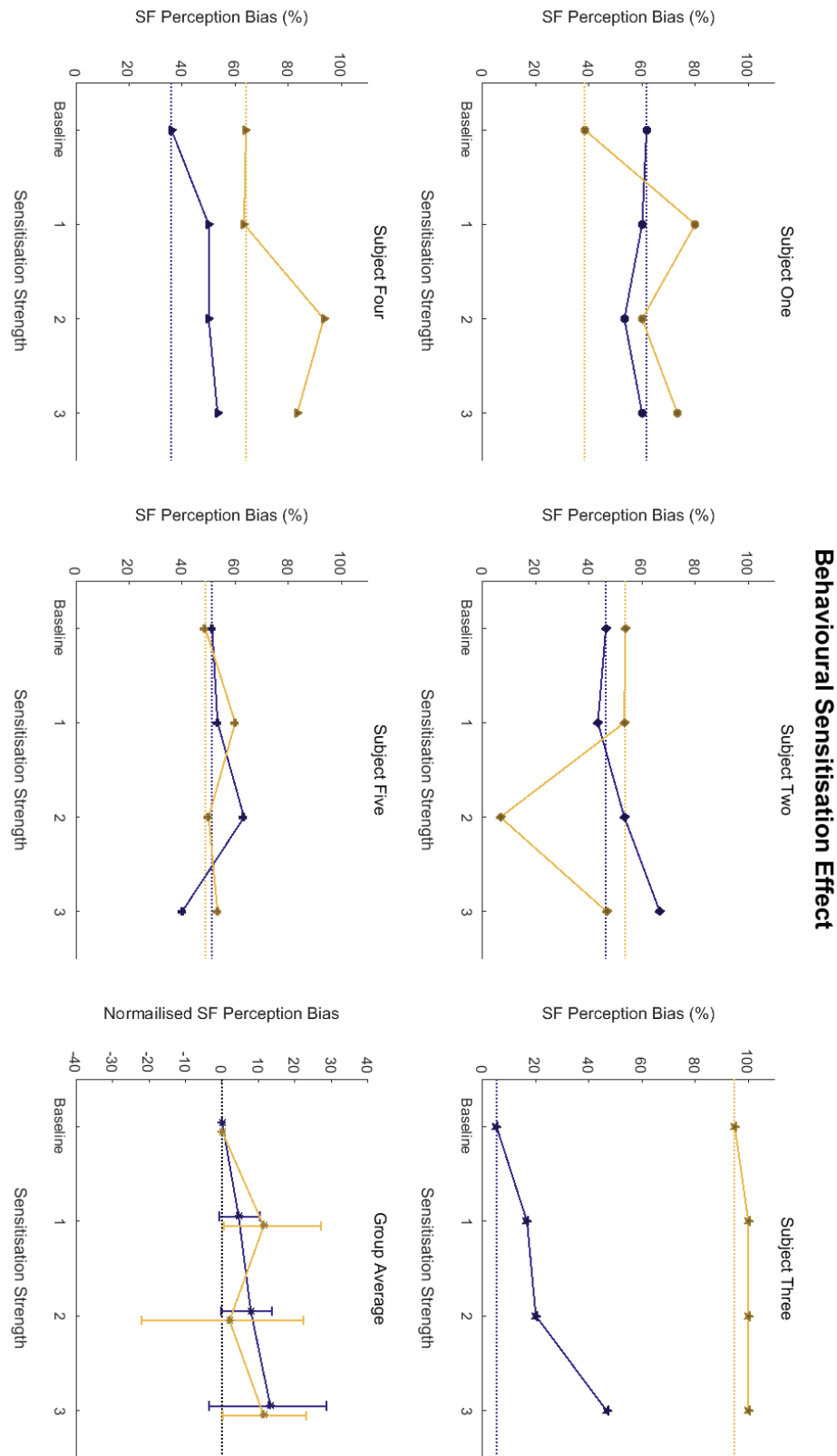
Figure 3-4 Behavioural results of sensitisation effect on perceptual bias. The first five panels show each subject's data. The y-axis represents classification scores in percentages. The x-axis shows sensitisation strength, where each number signifies sensitisation block number. Blue line shows LSF and yellow line HSF, where the horizontal lines represents each spatial frequency's baseline bias (acquired from un-sensitised hybrids). The sixth panel shows the group mean data normalised based on each subject's baseline bias. The group data was bootstrapped (with replacement, n=1000). Overall there seem to be indications of a positive relationship between sensitisation strength and perceptual bias.

At first glance, there seems to be a trend towards an increase in perception bias with sensitisation strength. When fitting a line-of-best-fit through the data of each subject, three out of five subjects have positive slopes for LSF, and four out of five subjects have positive HSF slopes. Bootstrapping the group means reveal a significantly positive slope in LSF perceptual bias with increasing sensitisation (slope: 4.31; p=0.043). The group mean slope of the HSF (slope: 2.4941) did not reach significance (p=0.19).

3.3.3.2 Classification Results

As shown in figure 3-5, we could not classify the category contained in the low frequency in any ROI (foveal: 49.1%, p=0.6648; peripheral: 51.0%, p=0.1193; far-peripheral: 51.4%, p=0.0762). However in the high frequency this was possible in foveal cortex, although the effect was small (foveal: 52.3%, p<0.0001; peripheral: 51.1%, p=0.1279; far-peripheral: 50.6%, p=0.2317).
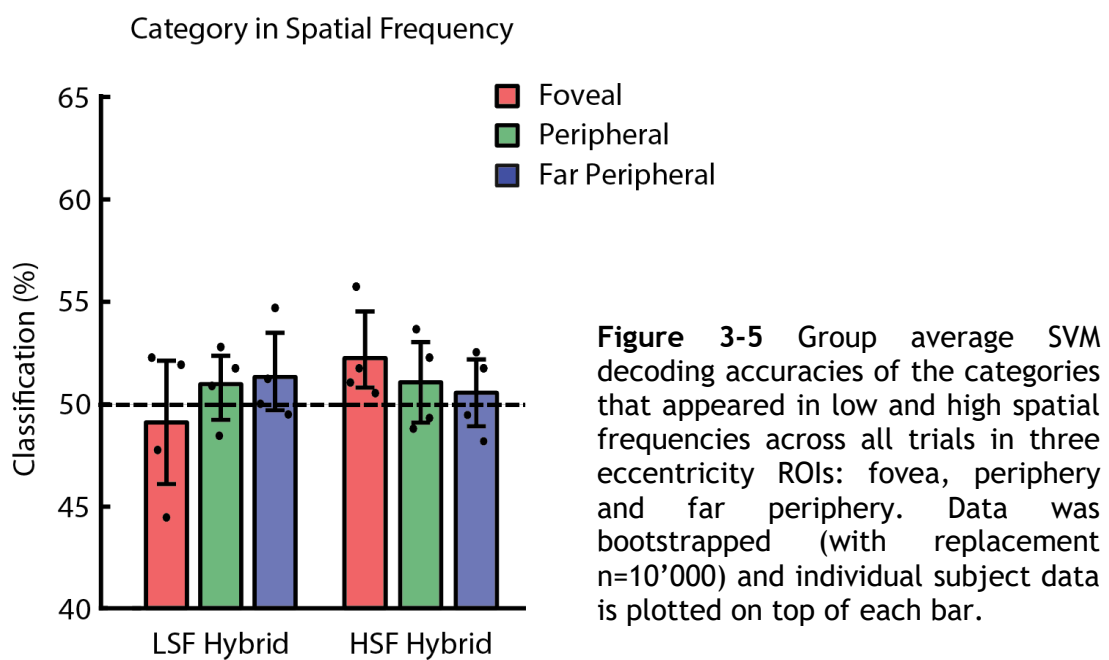
Figure 3-5 Group average SVM decoding accuracies of the categories that appeared in low and high spatial frequencies across all trials in three eccentricity ROIs: fovea, periphery and far periphery. Data was bootstrapped (with replacement n=10'000) and individual subject data is plotted on top of each bar.

Interestingly, classification of the reported category perception was significant in all ROIs (figure 3-6). This was the case when all hybrid trials were included (foveal: 63.4%, p<0.0001; peripheral: 62.4%, p<0.0001; far-peripheral: 62.7%, p<0.0001) or when only those hybrids were included where perception of low frequencies occurred (foveal: 61.3%, p<0.0001; peripheral: 61.2%, p<0.0001; far-peripheral: 61.5%, p<0.0001) and classification was particularly good when perception of high

frequencies occurred (foveal: 70.9%, p<0.0001; peripheral: 69.3%, p<0.0001; far-peripheral: 68.3%, p<0.0001).

This pattern of results was also largely true when we labelled all hybrids according to the category that was not reported by the subject (foveal: 59.5%, p<0.0001; peripheral: 60.5%, p<0.0001; far-peripheral: 59.6%, p<0.0001). This shows that, while not reported, the information was still available to the classifier. It is unclear whether the unreported category was actually perceived by the subject, since they were tasked with responding to whichever category seemed perceptually stronger. The analogous classifications for perception of low frequencies were lower and only significant in the peripheral ROI (foveal: 54.8%, p=0.0612; peripheral: 55.9%, p=0.0040; far-peripheral: 55.4%, p=0.062). Once again, particularly good classifications of high frequency content resulted even when this was not the reported category (foveal: 69.8%, p<0.0001; peripheral: 70.3%, p<0.0001; far-peripheral: 71.1%, p<0.0001).
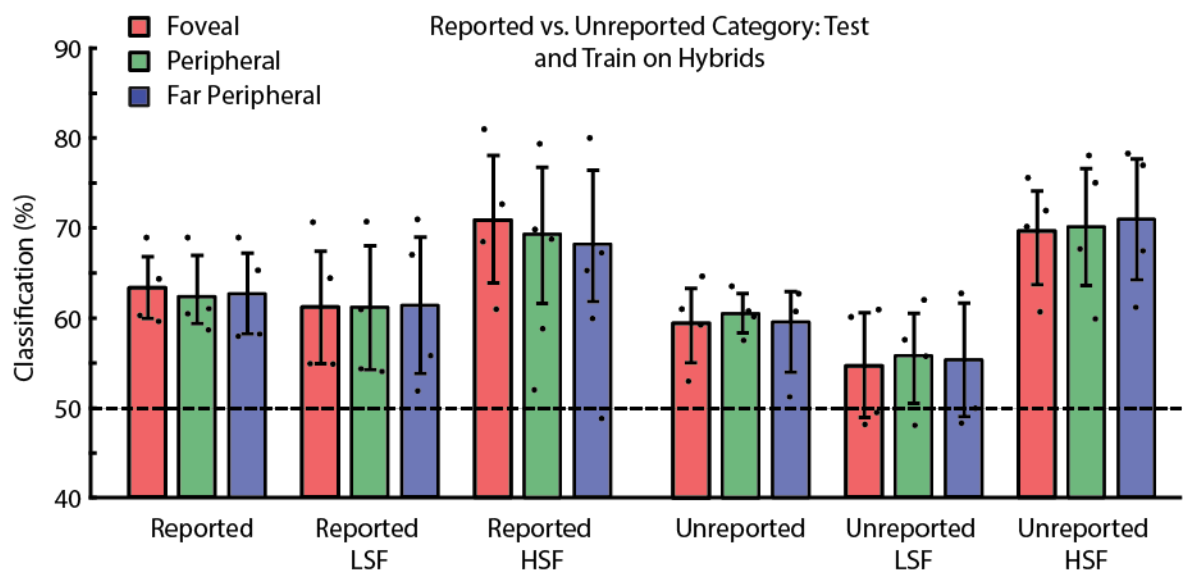
Figure 3-6 Group average SVM decoding accuracies of reported and unreported categories across all trials in three eccentricity ROIs: fovea, periphery and far periphery. Data was bootstrapped (with replacement n=10'000) and individual subject data is plotted on top of each bar.

In the control analysis to evaluate the possible contribution of motor responses to the classification of reported category (figure 3-7), we found that we could classify above chance which hand was used to respond (left vs. right: foveal:

64.4%, p<0.001; peripheral: 60.6%, p<0.001; far-peripheral: 62.5%, p<0.001), implying that motor responses may have played a role in our ability to classify in previous analyses. By studying figure 3-7, it can be seen that one subject in particular (denoted by the star symbol) had high classification scores for motor responses.
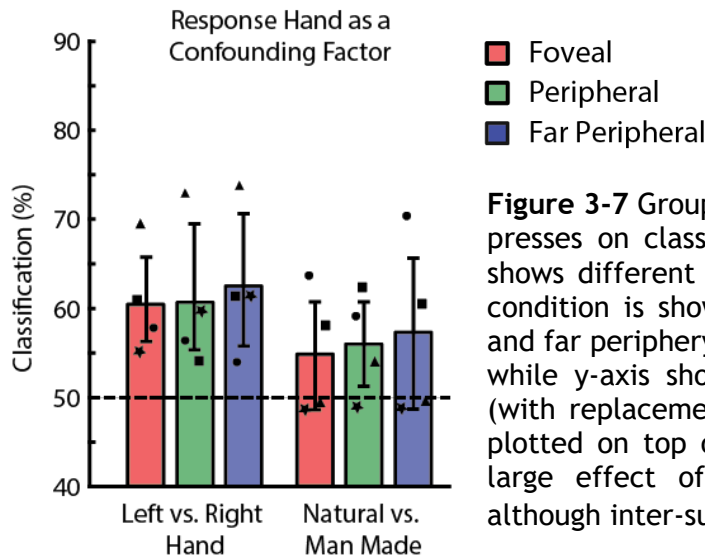
Figure 3-7 Group averages of influences of button-presses on classification results. Each bar colour shows different conditions (see legend) and each condition is shown in each ROI: fovea, periphery and far periphery. X-axis shows decoding accuracy, while y-axis shows ROIs. Data was bootstrapped (with replacement n=1000) and individual data is plotted on top of each bar. There seem to be a large effect of right vs. left hand condition, although inter-subject variability seems large.

3.4 Discussion

Our behavioural data showed a trend towards a linear relationship between sensitisation strength and perceptual bias. In other words, there were indications that with increased sensitisation the participant reported more frequent perception of the sensitised spatial frequency direction. However this effect was based on only five subjects (four of which were available for fMRI analysis) and very weak, with an average shift in bias of only a few percent. This hindered our ability to evaluate the effect of top down influences on spatial frequency processing in early visual cortex. Despite this, some positive findings emerged.

For the most part we were unable to classify hybrid categories labelled according to their low or high spatial frequency content. However, the reported category was readily available to the classifier as was the unreported category – particularly in the case of high spatial frequency content. These results seem inconsistent – why can we robustly classify the high spatial frequency content: 1) when it is reported, 2) when it is unreported but not 3) when we label this content in all trials, regardless of whether it was reported or not? Given that the motor

responses may have affected our data, it may be that the classifier was able to use that fact that a given unreported category was never associated with its mapped button response (but would have been associated with button responses to the other 3 categories) to achieve above chance classification. For example, when classifying unreported high spatial frequency occurrences of 'Mountain' and 'City', the consistent absence of a particular button press (e.g. for 'Mountain') might have resulted in a consistent difference in the neural pattern as compared to the unreported 'City' trials – which would occasionally have been associated with a button press for 'Mountain' when this category was being reported in the low spatial frequency.

If motor responses can be ignored, the result that classification of both reported and unreported categories is possible suggests that in early visual cortex information from both low and high spatial frequencies remain despite the bistable nature of the stimuli. However, since there were only 24 images at play and this experiment had 930 total trials, subjects quickly became familiar with all of the images. After the experiment subjects said that they often recognised both images in hybrid trials, and reported whichever seemed 'perceptually stronger' (as instructed). This is undesirable from an experimental standpoint. Another barrier to bistable perception in experiment 1 was that the spatial frequencies were Gaussian filtered. Such a filter is suboptimal in this case because it merely reduces, but does not eliminate, the power in various frequency bands – that is, it does not achieve a sharp spatial frequency cut-off. This latter point may go some way to explaining why we did not observe any differences in spatial frequency information among foveal, peripheral and far peripheral ROIs.

Considering the novelty of the paradigm in an fMRI setting and the low sample size, it is intriguing that some trends emerged. It is important to further underscore that the data and their generalisability needs to be interpreted with caution. Despite this, some interesting patterns were revealed which justify continuing to improve the paradigm to further explore the potential use of hybrid images in investigating cortical feedback. We sought to do this in experiment 2.

3.5 Experiment 2

3.5.1 Introduction

There were a number of shortcomings in experiment 1 which we attempted to address in experiment 2. The most obvious was to simply counter-balance the finger response mapping across runs to eliminate any motor confounds.

The sluggish nature of the BOLD signal makes it impractical to attempt sensitisation using many rapidly presented trials if individual trial estimates are needed. Therefore the main difference between experiments 1 and 2 was to abandon sensitisation as a manipulation. We still included sensitisation images, but these were now randomly interspersed with hybrid images throughout the run. We hereafter refer to sensitisation images as 'probes'. Our aim was to investigate the relation of neural patterns generated by probes and hybrid trials in which the spatial frequency perceived matched the spatial frequency of the probe.

In order to avoid recognition effects and therefore to enhance the bistable nature of the stimuli, in experiment 2 we used many more images - presenting each image only twice per experimental session. To enhance the bistable nature of hybrid stimuli, previous research has used a second-order Butterworth filter (for details, see Schyns & Oliva, 1994) to better separate the spatial into exclusively low and high spatial frequencies. We take the same approach in experiment 2 choosing the filter cut-offs such that there was a distance of 2 octaves between spatial frequency bands. This should help to avoid overlapping stimulation of the same neurons in early visual cortex by the different spatial frequency filtered images, as previous neurophysiological studies have shown that the average spatial frequency bandwidth of neurons in macaque V1 ranges from 1-1.5 octaves (De Valois, Albrecht, & Thorell, 1982).

We hypothesised that if only one of the two presented images in a hybrid can be processed at a time, the activation pattern in early visual cortex should only represent information about the perceived, but not the unperceived image. We were also interested in whether a hybrid image perceived in a given spatial frequency generated a similar pattern to a probe image of the same spatial

frequency. Such a correspondence would be evidence that the non-perceived image in the hybrid is akin to noise as is actually the case in the probe stimuli. Once again, we analysed the data in foveal, peripheral and far peripheral early visual cortex.

3.5.2 Methods

Unless otherwise stated, the details of experiment 2 were identical to experiment 1.

3.5.2.1 Subjects

Five healthy subjects (3 male; mean 24.8 years old, range = 21-26).

3.5.2.2 Stimuli and Apparatus

To avoid recognition of learned images, each hybrid and probe stimulus contained unique natural scene images selected from the SUN database (Xiao et al., 2010). For each category, 180 images were selected, resulting in a total of 180 unique hybrids, 90 LSF, and 90 HSF probes (therefore, each image appeared once in a hybrid stimulus and once as a probe stimulus). Images were cropped to 768´768 pixels, transformed into Fourier space and filtered in either HSF or LSF using a second-order Butterworth filter (for details, see Schyns & Oliva, 1994) with cut-off values of 1.35cpd and 5.4cpd for LSF and HSF, respectively. Thus, there was a distance of 2 octaves between spatial frequency bands. Throughout the experiment, a small white cross appeared at the centre of the screen. We changed the fixation cross so that it could be continuously displayed without obscuring the central image region. This change obviated the need to briefly remove the fixation from the display when the image appeared (as was done in experiment 1) which could conceivably interfere with foveal processing.

A. Hybrid: Mountain in HSF, Beach in LSF

B. Hybrid: Mountain in LSF, Beach in HSF
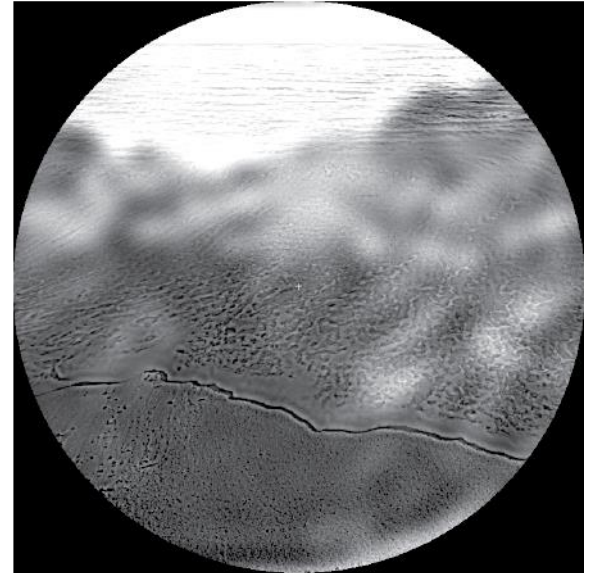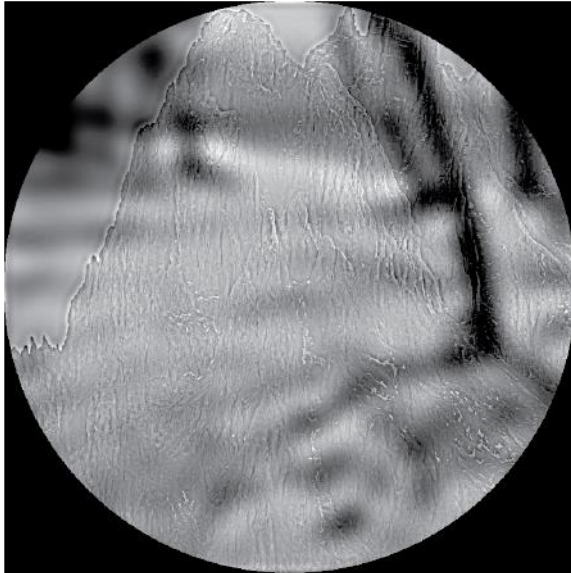
C.   Probe: Beach HSF
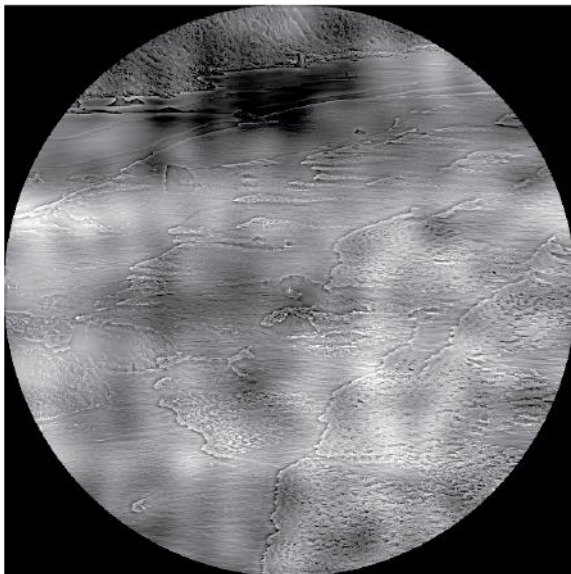
D.   Probe: Mountain LSF
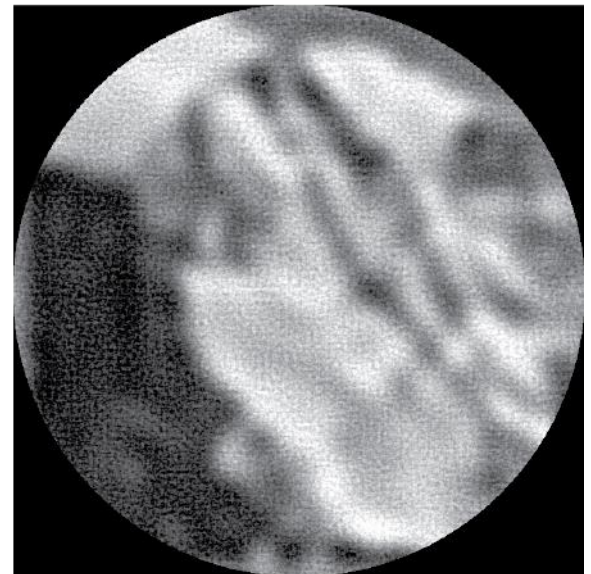
Figure 3-8 Examples of the experimental stimuli. **A & B.** Hybrid stimuli: A mountain and city scene combined. Panel A shows the mountain scene in high spatial frequency and city scene in low spatial frequency, panel B shows the same images but in opposite spatial frequencies. **C & D.** Probe stimuli: A forest scene in either high spatial frequency (panel C) or low spatial frequency (panel D) combined with a noise image of opposite spatial frequencies. Note that using the Butterworth filter produces a more bistable percept than does the Gaussian filter used in experiment 1 (figure 3-2).

3.5.2.3 Task and Procedure

Throughout the experiment, subjects fixated on a small white cross in the centre of the screen. Each experimental session consisted of six experimental runs, two retinotopic mapping runs, and an anatomical scan. In total, each session lasted approximately 1.5 hours. Experimental runs started and ended with a 20s baseline during which a grey screen was presented. During each experimental run, 30 hybrids, 15 LSF probes, and 15 HSF probes were presented in a randomised order for 120ms each with an inter-stimulus interval (ISI) of 5,880ms. Participants pressed one of three buttons on a button box in their right hand to indicate which of the three image categories they perceived. Button responses for the different categories were counter-balanced between runs, to avoid a potential motor response confound in early visual areas. Each experimental run lasted 6 minutes and 40 seconds.

3.5.2.4 Data Acquisition

Functional and anatomical MRI data was acquired using a 3 Tesla MRI system (Siemens Tim Trio) with a 32-channel head coil. For the functional scans an EPI sequence was used with the following parameters: 17 slices, aligned with the calcarine sulcus, gap thickness 0.3mm, TE-24ms, 395 volumes per run, and a resolution of $3mm^3$. We did not acquire a T2 weighted anatomical image.

3.5.2.5 Data Preprocessing

The functional data were aligned with the high resolution T1 weighted anatomical data.

3.5.2.6 Data Analysis

As in experiment 1, one option was to label all hybrid trials according to the category in either the low or high spatial frequency. We therefore attempted category classification of the low or high frequency content. We also attempted category classification of probe images (which contain a meaningful category in only one frequency).

Next, we attempted to classify the reported category. Given that each trial had a unique category exemplar (rather than 1 of a possible 6, as in experiment 1), our first step was to ensure that category classification was possible at all. We thus trained and tested an SVM classifier on all trials (hybrids and probes) labelled according to reported category. We also attempted the same analysis for trials in which low the spatial frequency or in which high spatial frequency image was reported (note that in the case of probe trials, there was only one perceptual option available for the subject).

We then analysed the hybrid stimuli separately from the probes. As in experiment 1, we attempted to classify all hybrid trials which were labelled according the reported or according to the unreported category. We also attempted the same analysis for trials in which low the spatial frequency or in which high spatial frequency image was reported.

We were next interested in whether a hybrid image perceived in a given spatial frequency generated a similar pattern to a probe image of the same spatial frequency. To test this, we trained an SVM classifier on probe images of a given spatial frequency and attempted to classify the image in the same frequency in hybrids trials, both in trials when that frequency was reported or not reported.

3.5.3 Results

3.5.3.1 Behavioural Results

We plotted behavioural responses to ensure participants were able to accurately perceive scene images in both high and low spatial frequencies across the different stimulus types. All participants were able to correctly classify images as one of the three stimulus categories (beach, forest, or mountain) with average accuracies of 79.8%, 74.7%, and 91.1% for LSF probes, HSF probes, and hybrids, respectively (figure 3-9 A). Accuracy was higher for the hybrid stimuli, presumably because chance level was 66% (as there are two images from three possible categories presented), as compared to 33% in the probes. The fact that accuracy is lower for probe trials suggests that subjects were not able to fully examine both high and low spatial frequency bands to determine the informative frequency and

the category contain therein. In the case of hybrids, either frequency will yield a category response. This suggests that the stimuli are likely more bistable than experiment 1 and lead to exclusive to perception in a single frequency. Although there was a slight bias across participants to report the HSF image of a hybrid (mean LSF bias = 42.1%; see shaded area in figure 3-9 B), all participants were able to report the LSF image in some trials and the HSF image in other trials. Interestingly, individual biases remained moderately stable across experimental runs (see individual lines in figure 3-9 B).
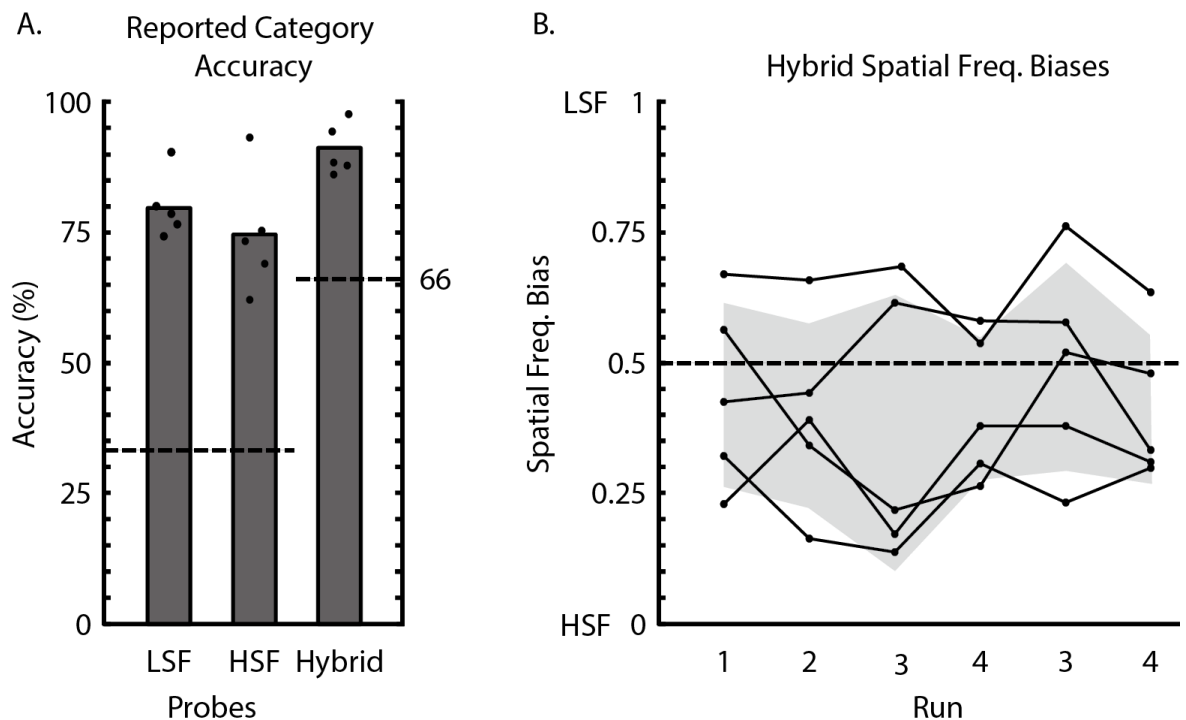
Figure 3-9 Behavioural results. **A.** Mean accuracy of reported category for low and high spatial frequency probes and hybrids. Bars show the mean accuracy across participants and black dots show individual accuracy rates across runs. Dashed lines indicate chancel level (33% and 66%, respectively). **B.** Inherent spatial frequency bias in hybrid trials. Lines show individual subjects across runs. Shaded area shows +/- 1 standard deviation around the group mean.

3.5.3.2 Classification Results

We emphasise at the outset that although we will note differences between ROIs and conditions, these are only trends seen by eye in the data and are not supported by statistical tests. We justify this approach only by the fact that we are limited to the data at hand and use it as a base for exploratory thought in

favour of simply concluding more data is needed to draw conclusions (while at the same time fully endorsing this latter point of view).

As shown in figure 3-10, we could not classify the category contained in any ROI in the low frequency probes (foveal: 46.4%, p=0.8403; peripheral: 50.6%, p=0.4078; far-peripheral: 50.9%, p=0.3705). This was also the case for high frequency probes (foveal: 51.0%, p=0.2972; peripheral: 54.5%, p=0.0758; far-peripheral: 52.3%, p=0.1475). For hybrid trials, we were able to classify the category contained in the low frequency the foveal and peripheral but not in far peripheral ROIs (foveal: 51.7%, p=0.0088; peripheral: 52.6%, p<0.0001; far-peripheral: 50.6%, p=0.2804). Interestingly, this was also the case in high frequency (foveal: 54.2%, p<0.0001; peripheral: 53.5%, p<0.0001; far-peripheral: 51.0%, p=0.3477). The failure to classify probes above chance is likely due to the fewer trial numbers (15 of each probe frequency per run) compared to hybrids (30 per run), as evidenced by the increased variability (figure 3-10). As such, we will not focus on probe trials in isolation. What is interesting is the Hybrid trials – although the errorbars overlap, the similarity in the fovea-periphery-far periphery gradient is noteworthy.
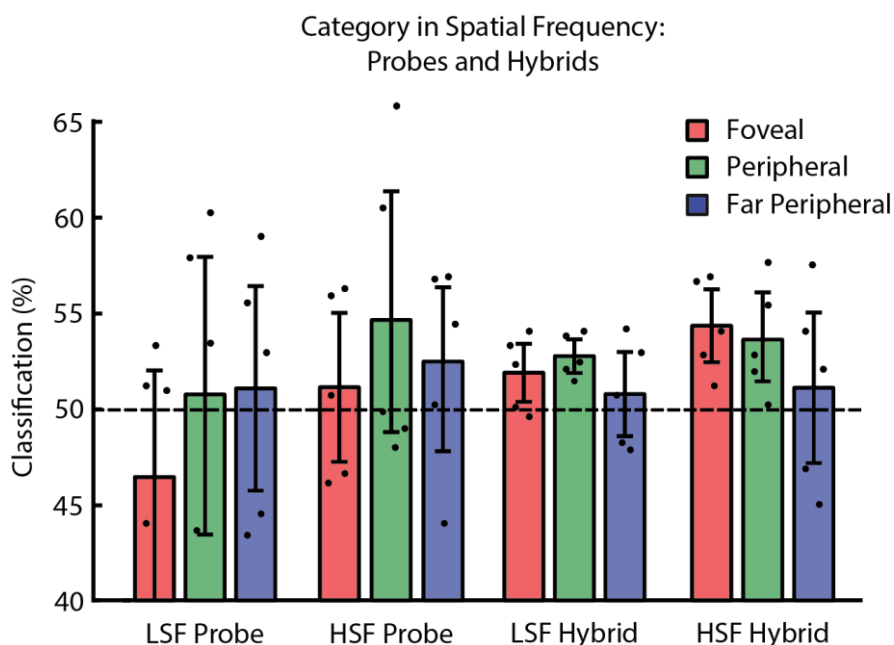
Figure 3-10 Group average SVM decoding accuracies of the categories that appeared in low and high spatial frequencies, in probes and in hybrids, in three eccentricity ROIs: fovea, periphery and far periphery. Data was bootstrapped (with replacement n=10'000) and individual subject data is plotted on top of each bar.

Next, we attempted to classify the reported category using all trials (hybrids and probes) to give maximum power to the SVM and thus show what accuracies we might expect in subsequent analyses. Classification of the reported category perception (figure 3-11) was significant in peripheral and particularly good in far peripheral ROIs when all trials were included (foveal: 50.5%, p=0.3684; peripheral: 53.3%, p=0.0002; far-peripheral: 55.7%, p<0.0001). The first thing to notice is the reversal of the fovea-periphery-far periphery gradient pattern observed earlier. Also Interesting, is that when only those hybrids where perception of low frequencies occurred we were additionally able to classify above chance in foveal cortex (foveal: 53.6%, p=0.0021; peripheral: 54.6%, p<0.0001; far-peripheral: 54.1%, p<0.0001) and the same pattern emerged in hybrid trials where perception of high frequencies occurred (foveal: 52.2%, p=0.0141; peripheral: 53.9%, p<0.0001; far-peripheral: 54.8%, p<0.0001). Increased classification accuracy in foveal cortex for the latter two comparisons is particularly surprising as there are about half as many trials to train the classifier in these cases which would decrease power as compared to using all trials.
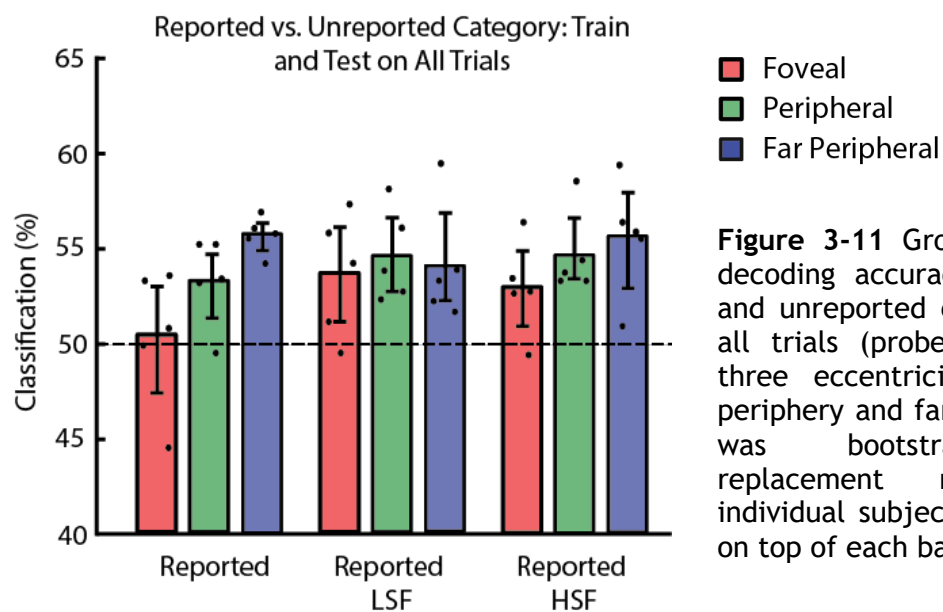
Figure 3-11 Group average SVM decoding accuracies of reported and unreported categories across all trials (probes & hybrids) in three eccentricity ROIs: fovea, periphery and far periphery. Data was bootstrapped (with replacement n=10'000) and individual subject data is plotted on top of each bar.

When running the above comparisons in hybrid trials only, we see a similar pattern of results, albeit weaker owing to less training data and also not corrected for multiple comparisons and should therefore be treated only as a starting point for cautious speculation (figure 3-12). Classification of the reported category perception was significant in the far peripheral ROIs when all hybrid trials were

included (foveal: 49.2%, p=0.8585; peripheral: 51.3%, p=0.218; far-peripheral: 53.0%, p=0.0214). The peripheral advantage remained when only those hybrids where perception of low frequencies occurred were used (foveal: 51.3%, p=0.2768; peripheral: 56.7%, p<0.0001; far-peripheral: 54.0%, p=0.0156) and a similar, but non-significant, trend appeared in hybrid trials where perception of high frequencies occurred (foveal: 48.6%, p=0.8373; peripheral: 48.5%, p=0.9076; far-peripheral: 52.4%, p=0.0862). It is interesting that we see a foveal advantage when classifying the category shown in a particular frequency, regardless of the frequency of the reported percept, but we see a peripheral advantage in the case of the reported percept. If confirmed with further data, and bolstered with appropriate corrections for multiple comparisons, this would represent an intriguing observation.
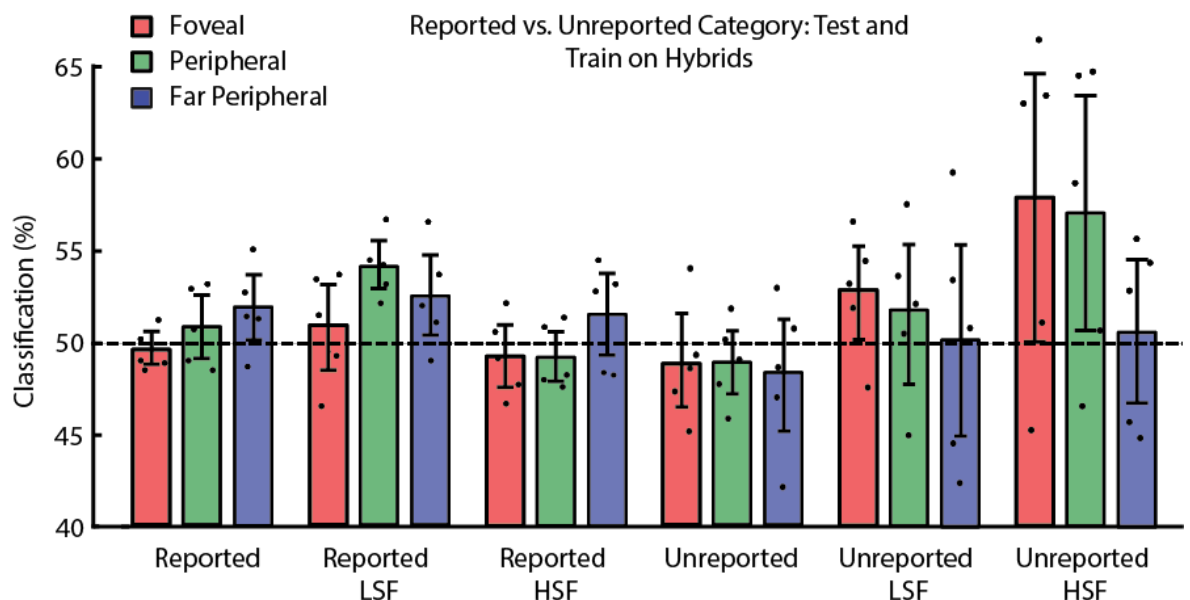
Figure 3-12 Group average SVM decoding accuracies of reported and unreported categories across all trials in three eccentricity ROIs: fovea, periphery and far periphery. Data was bootstrapped (with replacement n=10'000) and individual subject data is plotted on top of each bar.

Labelling hybrids according to the unreported category resulted generally in more variable classification accuracies across subjects, despite the number of trials in this analysis remaining the same. We will report these results, but not speculate about them with this additional source of uncertainty on top of all the aforementioned uncertainty. Using all hybrids, we were not able to classify above chance in any ROI (foveal: 48.9%, p=0.8125; peripheral: 48.9%, p=0.8769; far-

peripheral: 48.3%, p=0.8390). The analogous classifications for perception of low frequencies were significant in the foveal ROI (foveal: 52.8%, p=0.0254; peripheral: 51.7%, p=0.1758; far-peripheral: 50.1%, p=0.4815). For high frequency perception classifications, we were able to classify above chance in foveal and peripheral ROIs (foveal: 57.8%, p=0.0306; peripheral: 57.0%, p=0.0179; far-peripheral: 50.5%, p=0.3549), although the subjects were particularly variable.

Lastly we looked at whether probes of a given frequency could be used as a model to classify the category in the same frequency in hybrid trials both when that frequency was reported and when it was not reported (figure 3-13). Nearly all results were non-significant. We found that models trained on low frequency probes could classify categories reported in the low frequency hybrid image in the periphery (foveal: 49.4%, p=0.8113; peripheral: 52.9%, p<0.0003; far-peripheral: 50.9%, p=0.2930). Whereas, attempts to classify unreported low frequency hybrid categories failed (foveal: 50.8%, p=0.3651; peripheral: 49.6%, p=0.6355; far-peripheral: 47.3%, p=0.9998). Conversely, models trained on high frequency probes could classify categories reported in the high frequency hybrid image in the fovea (foveal: 53.5%, p<0.0001; peripheral: 51.4%, p=0.0999; far-peripheral: 49.5%, p=0.6105). Once again, classifying unreported high frequency hybrid categories failed (foveal: 46.8%, p=1.0000; peripheral: 47.0%, p=0.9257; far-peripheral: 49.6%, p=0.7899).
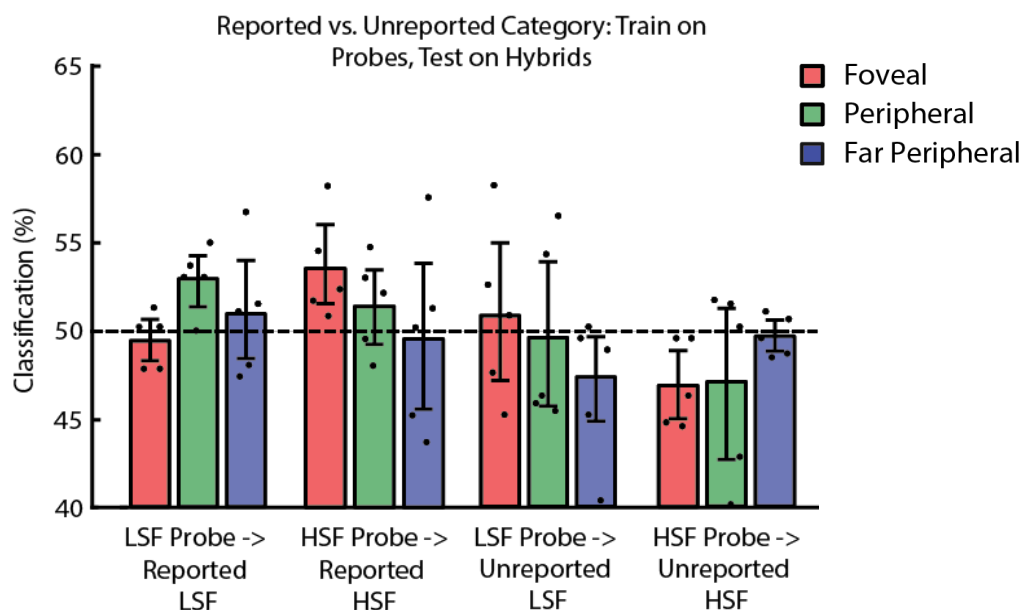

Figure 3-13 Group average SVM decoding accuracies reported and unreported hybrid categories, using models trained on low or high frequency probes, in three eccentricity ROIs: fovea, periphery and far periphery. Data was bootstrapped (with replacement n=10'000) and individual subject data is plotted on top of each bar.

3.5.4 Discussion

Experiment 2 was largely an attempt to improve on experiment 1. We conducted identical analyses as in experiment 1, in addition to some novel ones afforded by the probe trials. In experiment 2 classifying the category in low and high frequency channels in hybrid trials, regardless of which category was perceived, was successful in foveal and peripheral (but not in far peripheral) cortex. This contrasts with experiment 1, in which only high frequency information was available and then only in foveal cortex. The increased ability to classify in experiment 2 may have been due to a better separation of spatial frequency information as a result of the Butterworth filter.

When grouping hybrid and probes trials according to low or high frequency perception, we were additionally able to classify the category in far peripheral cortex. While the data presented here are too few to support any strong conclusions, we speculate that further support for the role of category perception per se in far peripheral cortex comes from analyses in which category perception was classified across all trials regardless of the perceived frequency. In this case we again saw an advantage for far peripheral but now also coupled with a lack of classification in foveal cortex – in contrast to classifying the category in low and high frequency channels regardless of category perception. That category could be classified in foveal cortex when grouping trials according to low or high frequency perception but not when pooling all trials could be taken to suggest that low and high frequency information is represented differently in foveal cortex whilst in far peripheral cortex a more generalised/courser category representation occurs (since perceived category classification was successful in all cases).

The trend for a far peripheral advantage for classifying perceived category described above was also apparent when only hybrid trials (but not probes) were considered. In addition, as in experiment 1, we also looked at classification of the unreported category. Unlike in experiment 1, in which both the reported and unreported category was readily available to the classifier in all ROIs, in experiment 2 only the reported category was available when pooling all hybrid trials and then only in the far peripheral ROI. Again we see an advantage for far peripheral cortex when classifying perceived category.

The last analysis was to train a model on low or high frequency probes and attempt to classify the category in low or high frequency channels in hybrid trials. Nearly all results were non-significant and so we urge extreme caution in the following discussion. We tried this approach both on reported and unreported hybrid categories. While no information was available for unreported categories, we found that classification of reported low and high frequency categories was possible in peripheral and foveal cortex, respectively. It is interesting that the established far peripheral advantage for perceived category classification was not present in this analysis. One speculative explanation is that some process occurs in far peripheral cortex for hybrid but not probe images (which then hinders cross-classification). Perhaps the presence of two images in hybrid trials induces some form of suppression of the non-perceived category. Since this process would not occur in probe images (since only a single image exists) the neural patterns in far peripheral cortex are dissimilar. The absence of such a suppressive process in foveal and peripheral cortex during hybrid trials would perhaps result in a similar neural pattern to corresponding probe trials. This interpretation must be taken with caution both because of the small number of subjects and also the fact that regular classification of probe trials failed (figure 3-10).

In experiment 2 we overcame many of the shortcomings of experiment 1. As such, the trends regarding an advantage in classifying perceived category from far peripheral cortex and in classifying category regardless of perception from foveal cortex are intriguing. However, before any of the results presented here can be taken seriously, much more data would be required to confirm them.

3.6 Chapter Discussion

We presented hybrid images to subjects while recording fMRI in two experiments. In experiment 1 we found that we could classify the reported and the unreported category, regardless of the frequency in which it was perceived. This could be interpreted as evidence that image content from both frequency channels was present in early visual cortex and that only in higher areas does the visual system select one of the images to enter conscious awareness. However such an interpretation is problematic due to various shortcomings in the design and in the stimuli used for experiment 1. It seems probable from self-reports that subjects

consciously perceived both images in our hybrid stimuli and thus classification of both images would be expected.

In experiment 2, we addressed these shortcomings and obtained different results. We found that the perceived category was generally classifiable in far peripheral early visual cortex. Conversely, foveal cortex seemed to represent information independently of category perception. While based on only five subjects, these results are intriguing although further data collection would be needed to confirm them.

The absence of category information in one spatial frequency channel (i.e. a failure to classify the unreported image) can be interpreted in one of two ways. First, it could be an indication that feedback information was sent to early visual areas and affected the feedforward information which ought to represent the physically present input from both spatial frequency channels. Second, it could be that on a given hybrid trial that some kind of biased competition-like process occurred within early visual cortex and was resolved without any feedback influence, resulting in only one of the two possible feedforward representations becoming fully established and later fed-forward to higher areas. In experiment 2, since the subject was not instructed or induced to focus on a particular spatial frequency channel, and since all subjects perceived sometimes low and sometimes high frequencies, there need not have been a top down driving influence. However, we do not see evidence of such a biased competition process in foveal cortex – since we were able to classify both low and high frequency channel information, regardless of which was perceived. This suggests that the selection of a particular image/frequency occurred later in the cortical hierarchy, which leaves room for a feedback interpretation for our ability to classify reported category in the far periphery.

This raises the question of why only far peripheral cortex would be subject to such top down influences. It is especially interesting in light of the results from chapter 2 in which we repeatedly found evidence of scene related feedback to both foveal and peripheral cortex. In chapter 2 we suggested that "[...] retinotopic biases throughout the visual hierarchy provide an organisational scheme for segregated cortical feedback of information about distinct higher level stimuli." The results regarding the far peripheral cortex are more in line with the peripheral bias of

high level scene areas. One reason for the discrepancy could be that in experiment 2, each image only appeared twice in an entire experiment for just 120ms and as such only an overall gist perception would be possible – and even that in the presence of unrelated information in the other frequency channel. This kind of gross scene analysis might have been more in keeping with the capability of far peripheral cortex. The repetition of the same scene many times over the 12s block in the experiments of chapter 2 might have rendered foveal processing possible and even more reliable (Müller, Metha, Krauskopf, & Lennie, 1999).

We originally hypothesised that low and high frequency information might be preferentially fed-back to peripheral and foveal cortex, respectively. The only hint of such an effect is that low and high frequency probes allowed classification of reported, but not of unreported, low and high frequency categories in peripheral and foveal cortex, respectively. This hypothesis is not incompatible with the idea of eccentricity biases throughout the cortical as an organising principle for feedback. Both ideas may be true to some extent, and disentangling them is difficult with the current data given the inherent correlation between eccentricity and spatial frequency information in visual cortex and must remain a problem to be solved in future work.

If the trends observed can be confirmed with more data, there are several possibilities for future work that would be interesting to explore. One idea comes from the work in chapter 2; it would be interesting to create hybrid images using object rather than scene stimuli. Might this reverse the current findings such that the perceived object would be represented in foveal cortex, while both objects would remain in far peripheral cortex? Another approach would be to add an occluder to the hybrid stimuli – this should allow an examination of the feedback signals in isolation as well as in interaction with feedforward signals in non-occluded regions.

Other paradigm modifications could be made to more explicitly involve top down processes. An element of the original behavioural work with hybrid stimuli that were not able to investigate here is a top down influence designed to actively bias frequency perception. While sensitisation is difficult in an fMRI paradigm, some previous work required subjects to consciously attempt to 'tune in' to a given frequency (Sowden et al., 2003). Such a manipulation would be easier to

incorporate in an fMRI paradigm and (i.e. by a simple auditory cue as in Sowden et al., 2003) and might prove more reliable on a trial by trial basis. This might be a more promising approach to take in future work.

Different methods is yet another way to understand better the processes involved. Using ultra-high field fMRI would allow depth dependent classification. It would be exciting to find that in mid-layers (associated with feedforward input) information about categories in both frequency channels is readily available whereas information in superficial and/or deep layers (associated with feedback input) contained only the perceived category. Such a result would strongly argue for a feedback component in the perception of bistable hybrid images.

To look at sensitisation, perhaps fMRI is not the ideal imaging method. However, it ought to be feasible to design such a paradigm for use with imaging techniques with higher temporal resolution such as EEG and MEG. With such temporal resolution, it should be possible to capture the trial-by-trial changes in neural activity that may be informative in understanding which percept the observer is experiencing. Particularly interesting is the potential to measure pre-stimulus activity - if we could demonstrate systematic changes in baseline activity, brought about by sensitisation procedures, that predicted hybrid perception then this would be indicative of top-down influences. It might also be possible to track whether initially both images in a hybrid stimulus are represented and later one is extinguished or if there is only ever one image represented (similar to Scholte et al., 2008). This was something we could not easily determine in the present study.

One of the theoretical difficulties, discussed in this chapter, in studying top down processing in visual cortex using visual stimuli is the difficulty in disentangling feedback from lateral mechanisms. In the next chapter, we use non-visual stimuli to induce feedback in the visual system.

4 Chapter 4: High-resolution (7T) fMRI reveals auditory and imagery information across cortical depths in non-stimulated visual cortex

4.1 Abstract

Vetter, Smith, and Muckli (2014) showed that complex auditory scenes played to blindfolded subjects can be classified from V1, V2 and V3 using multivariate techniques. This raises questions about what the information represents and its relation to other non-feedforward input to visual cortex. Here, we use high-resolution 7T functional brain imaging and multivariate classifiers to study auditory information and mental imagery in the early visual cortex of blindfolded subjects. The high resolution allowed us to look at activity patterns in different depths through the grey matter.

4.2 Chapter Introduction

One of the challenges in studying feedback in early visual cortex is knowing where the feedback originates from. Since there are many visual areas in the cortical hierarchy, any one of them could plausibly be contributing to our observed effects. Similarly, while in chapter 2 we took measures to avoid lateral connections as a mechanism by which our results could be explained, we have no easy way of ruling this out conclusively. The reason for the two issues is the same: the feedforward input arrives into the same visual area(s) in which we wish to detect feedback information. One way of avoiding this problem is to study cross-modal feedback effects. That is, present a stimulus input to one modality and investigate stimulus related activity in a cortical system that receives feedforward input exclusively from a different modality. This is the approach we took in this chapter. We used naturalistic real-world auditory scene stimuli, which aided comparisons with the naturalist scene stimuli used in chapters 2 and 3. Also, complex high level stimuli should engage a larger number of high level areas (compared to simple stimuli)

that may feedback to early visual cortex. Therefore, using complex stimuli should maximise the chances of detecting feedback signals.

We replicated an experiment that our lab previously conducted at 3T (Vetter et al., 2014), which was designed to investigate auditory and imagery feedback in early visual cortex. The novelty in the replication was that we recorded data at a field strength of 7T (which we will occasionally refer to as 'ultra-high field'). This allowed us to achieve a higher functional resolution than is possible at 3T, which allowed us to look at activity patterns in different depths through the grey matter as was done by Muckli et al. (2015). This is important because, as was outlined in the general introduction, feedforward and feedback connections terminate with distinct laminar profiles.

Mental imagery acted as a control condition in this experiment, but is an attractive way to studying feedback in the visual system in its own right because it is entirely internally generated. As such, it provides an ecologically valid situation of feedback/top down signals in the absence of any feedforward driving input (e.g. most people presumably engage in some form mental imagery with closed eyes before going to sleep). As in with the occluded paradigm used in the first chapter, the absence of informative feedforward input is helpful in isolating feedback signals.

4.2.1 Introduction

The stimulus properties that drive occipital cortex have been extensively studied. The most basic feature common to almost all of these stimuli is that they are presented in the visual modality. This statement is not controversial and yet however it ignores the important observation that non-visual stimuli can also drive responses in early visual cortex. This fact has been known for over five decades and yet we still do not have a clear understanding of the information conveyed or the adaptive function of such signals (L. Petro, Paton, & Muckli, 2017).

Morrell (1972) found that over 40% of visually responsive V1 cells recorded in cats also respond to auditory stimuli. Moreover these cells responded selectively to sounds according to their displacement in the horizontal direction and this

location is predicted by the horizontal coordinate of their visual receptive field. This result were replicated by (Fishman & Michael, 1973).

Clavagnier, Falchier, and Kennedy (2004) demonstrated using retrograde tracers the existence of direct long distance feedback connections to V1 originating in primary auditory cortex (A1) and high higher multisensory areas (superior temporal polysensory or 'STP') in monkeys (a finding possibly unique to primates). This forces us to consider the functional role of auditory signals arriving into early visual cortex. One plausible idea is that such signals serve to orient gaze and/or visual attention to objects of interest in the world. Primates rely heavily on their visual systems, with about 60% of the cortex primarily responsive to visual stimuli. Therefore, information about a potentially important event that arrives through the medium of sound will likely soon be processed visually. Sound often might be the first input relating a feature of the external world because the external source of the sound may be located outside the visual field of view of the organism. In such situations, as the animal must orientate its gaze to bring the source of the sound into view. In such cases, the first part of the visual field that will receive input will be the far periphery.

Falchier, Clavagnier, Barone, and Kennedy (2002) showed using retrograde tracers in monkeys that feedback connections to V1 from primary and higher auditory cortex as well as from STP project mainly to peripheral cortex with very few connections made to foveal cortex. Similar results were found by Kathleen S Rockland and Ojima (2003) in V1 and even denser in V2. As would be expected the feedback connections originated from infra-granular layer 6. Such results are understandable if the purpose of auditory feedback is to aid orientating gaze or visual attention. However this does not easily explain reports that individuals show stronger effects of auditory stimulation in visual cortices particularly in peripheral visual cortex (Burton, 2003).

Vetter et al. (2014) blindfolded normal sighted subjects passively listened to one of three complex natural scene sounds while recording fMRI. They found that classification of the sounds was strongest in far peripheral visual cortex and not possible in foveal cortex. While the result is in line with the peripheral bias of auditory feedback projections, it is hard to explain with a simple gaze/attention orientation hypothesis. Vetter et al. also investigated mental imagery – in which

subjects imagined the sounds and the corresponding natural scenes. Interestingly, classification of mental imagery content was strongest in foveal cortex, and not possible in far peripheral cortex – in stark contrast to sound content. This different eccentricity profile argues against involuntary mental imagery as an explanation for the ability to classify auditory information.

We replicated the experiment by Vetter et al. (2014) using high resolution 7T fMRI, which allowed us to look at activity patterns in different depths. A similar approach was taken by Muckli et al. (2015) using an occluder paradigm (similar to the experiments in chapter 2). The key finding was that feedback signals were detected in supra-granular depths – where feedback connections are known to terminate whereas feedforward signals could additionally be detected in middle depths – where feedforward inputs arrive. However, feedback activity in infra-granular depths has been reported in 7T fMRI by Kok, Bains, van Mourik, Norris, and de Lange (2016) using a visual illusion paradigm. While long distance feedback connections are known to terminate in supra-granular layers, some feedback connections from nearby cortical areas terminate in infra-granular layers (Markov et al., 2013; Markov et al., 2014). It may be that the different stimuli engaged different higher areas leading to divergent results. These studies demonstrate the feasibility and usefulness of using 7T fMRI to investigate feedback processes across cortical depths, but also highlight the complexity of the laminar connectivity profile.

Given the divergent results of Muckli et al. (2015) and Kok et al. (2016), the question of which cortical depths we should expect to see auditory and imagery feedback in is difficult. One study in mice found auditory feedback to visual cortex drove activity in infra-granular layers which in turn inhibited supra-granular pyramidal cells (Iurilli et al., 2012). This further highlights the difficulty in making strong predictions, as activity in on layer can drive activity in another layer. Therefore, the only prediction we can make for feedback information profiles across cortical depths is that we should not detect information in middle depths since these depths are driven by feedforward stimulation. Following Vetter et al. (2014), we also predicted that auditory information would be found in far peripheral but not foveal regions and that the opposite would be true of imagery information.

4.2.2 Methods

The experimental stimuli, task and procedure reported here are identical to those used in Vetter et al. (2014).

4.2.2.1 Subjects

Four healthy subjects (1 male; 20-26 years old) were recruited and measured at the Maastricht brain imaging Centre (Maastricht, Netherlands). All subjects had normal hearing and no history of brain damage. Each subject completed three consent forms to ensure that they understood of the experimental conditions and that they met the safety criteria for fMRI scanning according to the Ethical Committees at 1) of the College of Science and Engineering, University of Glasgow and 2) the Faculty of Psychology and Neuroscience, University of Maastricht.

4.2.2.2 Stimuli and Apparatus

Subjects wore a blindfold to prevent stimulation of the visual cortex. The sound stimuli were delivered binaurally through fMRI-compatible earphones (Sensimetrics Corporation). For each subject, the sound levels were adjusted to ensure subject comfort and that the sounds were audible above the sound of the scanner. The experiment was programmed and presented using Presentation (Version 16.5, Neurobehavioural Systems). We used the same three natural sound stimuli as was used in Vetter et al. (2014). The sounds were normalised for amplitude, consisting of traffic noise (a busy road with cars and motorbikes), a forest scene (birds singing and a stream) and of a crowded scene (people talking without clear semantic information) and were downloaded from www.soundsnap.com and cut to 12s.

4.2.2.3 Task and Procedure

Subjects wore a blindfold, were instructed to keep their eyes closed at all times and room lights were switched off. Subjects completed 8 functional runs. During odd runs we presented natural sound stimuli. The three natural sound conditions were each presented for 12s and repeated 6 times for a total of 18 blocks per run. A 12s baseline followed every block. During even runs, subjects were instructed

to imagine the scenes using any modality(s). The numbers of repetitions and their length was the same as in the auditory runs Subjects were instructed to imagine by a 1s verbal cue ("traffic", "forest" or "people") followed by a 12s imagery period during which no sounds were played. Termination of the 12s imagery periods was indicated to the subjects by a 1s long beep. A 12s baseline followed every block. In auditory and imagery runs, the condition order was pseudo-randomised with the constraint that no condition could be repeated back-to-back.

4.2.2.4 Data Acquisition

Functional and anatomical MRI data was acquired on the 7-T scanner (Siemens) in Maastricht (The Netherlands) with a 32-channel head coil (Nova Medical Inc, USA). For the non-multiband functional scans (1 subject) an EPI sequence was used with the following parameters: 31 slices, aligned perpendicular to the calcarine sulcus, gap thickness 0mm, TR-2s, TE-25ms, 228 volumes for auditory runs and 245 volumes for imagery runs (polar angle mapping = 340 volumes, eccentricity mapping = 283 volumes), a FOV of 128mm, flip angle of 85° and a resolution of $0.8mm^3$. For the multiband functional scans (3 subjects) an echo-planar imaging sequence was used with the following parameters: 56 slices, aligned perpendicular to the calcarine sulcus, gap thickness 0mm, TR-2s, TE-25ms, 226 volumes for auditory runs and 243 volumes for imagery runs (polar angle mapping = 396 volumes, eccentricity mapping = 288 volumes), a FOV of 128mm, a multiband acceleration factor of 2, flip angle of 75° and a resolution of $0.8mm^3$. All functional scans were preceded by 5 volumes with reversed phase-encoding direction (posterior to anterior) in order to allow estimation and correction of geometrical distortions due to magnetic susceptibility gradients (Andersson, Skare, & Ashburner, 2003). We also acquired three anatomical scans, all with a resolution of $0.6mm^3$: T1-weighted Anatomical, a proton density-weighted contrast and a T2*-weighted sequence.

4.2.2.5 Data preprocessing

The functional and anatomical data were preprocessed using BrainVoyager QX 2.8 (Brain Innovation, Maastricht, The Netherlands). We divided the T1 by the PD to correct for inhomogeneity biases. The first two volumes of each functional run were discarded to avoid saturation effects. The functional data for each run were

corrected for slice acquisition time and head movements. Linear and low frequency drifts in the data were removed. For each run, we estimated the degree of spatial distortion due to phase encoding direction by comparing the 5 volumes in which the phase encoding direction was reversed with the volumes in the run itself and corrected the data accordingly (for a full procedural description, see Andersson et al., 2003). The functional data were aligned to each other, then to high resolution the T2* anatomical scan and transformed into ACPC space.

4.2.2.6 ROI Definitions

For each subject, the functional data were projected onto the cortical surface. For the subject without multiband scans we only had partial functional coverage of V3. For those subjects with multiband acquisitions, we had enough functional coverage to collect data in V1, V2 and V3. We were able to collect retinotopic mapping data which allowed us to map the early visual areas in three subjects (one multiband subjects being the exception). The ROIs were defined (figure 4-1 A) by taking eccentricity bands across V1, V2 and V3 approximately corresponding to between 1-3°, 3-6° and 6-10+°. In the subject without retinotopic mapping, we manually estimated the regions from cortical folding.

A)   Early Visual Cortex ROIs

B)   Cortical depth ROIs in V1

Anterior

Posterior

- ■ Foveal
- ■ Peripheral
- ■ Far Peripheral

| 90 | 74 | 58 | 42 | 26 | 10 |

Cortical depth (%)

Figure 4-1 A. Example of ROI definitions in the right hemisphere. **B.** Example of 3 depth ROIs in both hemispheres in V1 (depths shown = 90%, 58%, 10%).
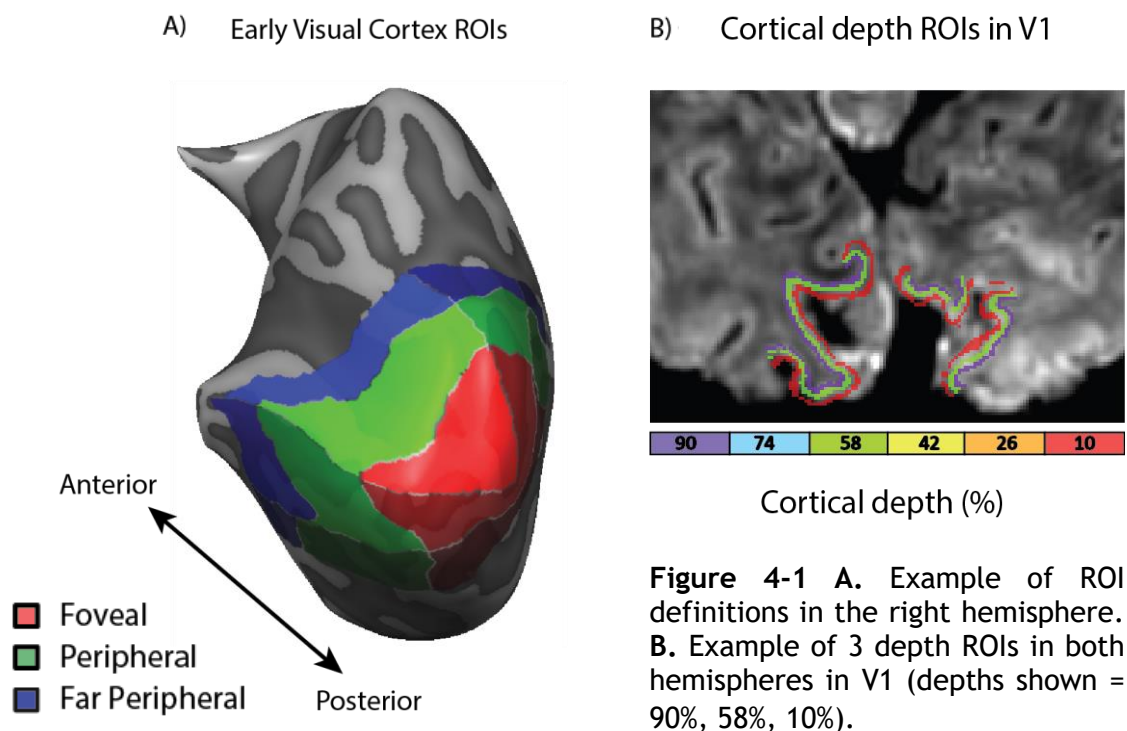
For each of the above described ROIs, we labelled the volumetric functional voxels according to 6 evenly spaced cortical depths (figure 4-1 B). This relied on first estimating the grids as vertices using a cortical thickness map derived from the T1 anatomical scan (voxels could be labelled by more than one depth). These vertices were then projected into volume space to label the voxels.

4.2.2.7 Data Analysis

For each Subject, a GLM was used to estimate each voxel's HRF amplitude for each block. To classify an auditory scene, an SVM classifier was used to create a discriminating function for the instances of the three auditory scenes presented, based on the associated multivariate voxel response patterns from 3 runs ("Training"). The discriminating function was then used to classify the scene instances presented during the remaining run ("Testing"). This procedure was repeated, each time using a different run as the "test run" (i.e. 'leave one run out' cross-validation procedure). The same procedure was used for classifying imagery conditions. To assess the overall localisation of auditory and imagery information in early visual cortex, we ran SVMs in ROIs which combined all layers. We then ran SVMs at six individual depths.

To assess the significance of group level effects, we bootstrapped (with replacement) the mean group classification value 10'000 times and if 95% of these values exceed chance level (33%) the classification was considered significant. To assess significance at the individual subject level in each depth ROI we employed the method described in Etzel and Braver (2013) who refer to it as the 'fold-wise scheme' and is commonly used for fMRI analyses. We trained and tested the SVM classifier 1000 times – each time permuting the training labels of the trials used to train the SVM on each cross-validation fold, thus precluding the opportunity to build a reliable discriminant function for the testing phase – except by pure chance – resulting in a chance level distribution. If the original, non-permuted mean classification accuracy falls outside this distribution (in the above chance direction), at an alpha of $p < 0.05$, then the classification is considered successful for the individual subject.

4.2.3 Results

We first looked at the SVM results from ROIs that combined depths in V1, V2 and V3 and which were subdivided into foveal, peripheral and far peripheral regions. Figure 4-2 A shows the results of classifying auditory conditions. We could classify significantly above chance in nearly every ROI. Specifically, classification was significant in all eccentricities in V1 (foveal: %41.7, $p<0.0001$; peripheral: 42%, $p<0.0001$; far-peripheral: 40%, $p<0.0001$) and V2 (foveal: %41.4, $p<0.0001$; peripheral: 42.3%, $p<0.0001$; far-peripheral: 39.9%, $p<0.0001$). Moreover, in each case the individual subjects were also above chance (with the exception of one subject in the V2 far peripheral ROI). In V3, the results were significantly above chance in foveal and peripheral, but not far peripheral ROIs (foveal: %40.3, $p=0.0197$; peripheral: 41.4%, $p<0.0001$; far-peripheral: 37.4%, $p=0.0724$). These results are in contrast to results from Vetter et al. (2014), who found that auditory information was localised to the far periphery, and strongest in V2.

Figure 4-2 B shows the results of classifying the imagery conditions in the same ROIs as described above. The classification of imagery information was not as uniformly above chance as the auditory classifications. Specifically, we were able to classify significantly above chance in all V1 ROIs (foveal: %37.1, $p<0.0001$; peripheral: 43.1%, $p<0.0001$; far-peripheral: 39.8%, $p<0.0001$), and in foveal and far peripheral V2 ROIs (foveal: 41.1%, $p<0.0001$; peripheral: 36.1%, $p=0.0587$; far-peripheral: 42.7%, $p<0.0001$), and in peripheral and far peripheral V3 ROIs (foveal: 34.8%, $p=0.3353$; peripheral: 38.5%, $p=0.0172$; far-peripheral: 39.2%, $p<0.0001$).

We decided to look at cortical depths in V1 but not in V2 and V3 ROIs. The reasons for this decision were both theoretical and pragmatic. From a theoretical standpoint, V1 is the most studied visual area (Olshausen & Field, 2005) and is thus an attractive starting point. From a pragmatic standpoint, V1 is the only ROI our lab has previously looked into with ultra-high field fMRI and it is also the only ROI in this experiment in which all subjects classified above chance in both auditory and imagery conditions. Last, as constructing cortical layer segmentation over large areas of cortex takes an excessive amount of time the aim is to do the same for V2 and V3 at a later point. At present we have done this for V1.
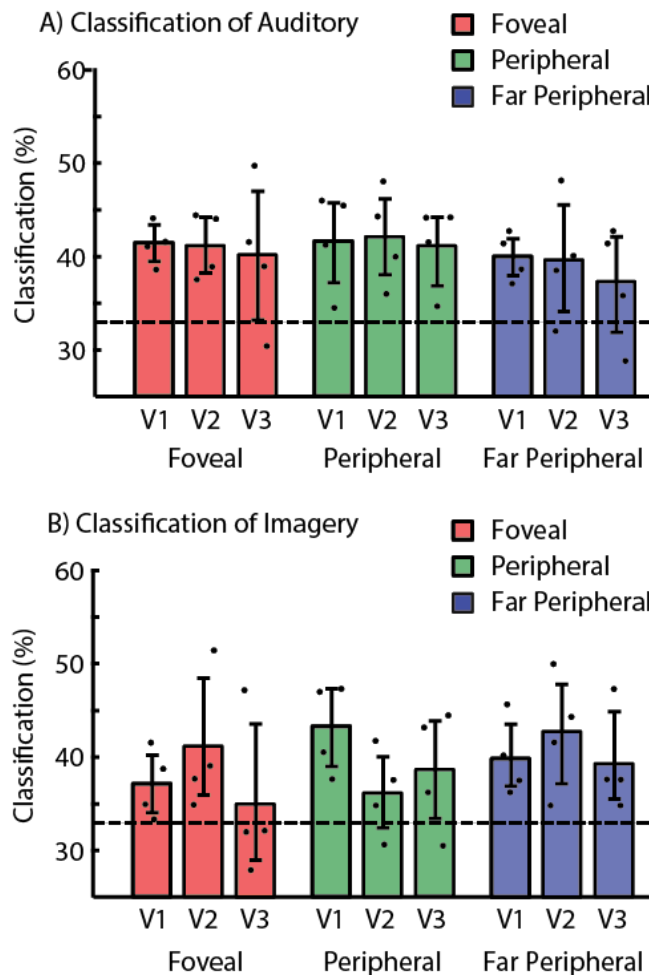
Figure 4-2 Group average SVM decoding accuracies in ROIs combining cortical depths. **A.** Auditory conditions. **B.** Mental imagery conditions. Data was bootstrapped (with replacement n=10'000) and individual subject data is plotted on top of each bar.

Given that we did not observe a difference in between foveal, peripheral and far peripheral V1 ROIs (or in V2 and V3 ROIs), we decided to collapse these V1 ROIs in order to give the SVM classifier the maximum number of V1 voxels available for each depth. Figure 4-3 shows the SVM classification in each depth. In reporting the statistics below, we refer to the most and least superficial depth ROIs as 'depth 1' and 'depth 6', respectively (with intervening depths labelled 5 through 2). Individual profiles, shown in Figure 4-3 A and B, were not as consistent as in Muckli et al. (2015). While two of the subjects showed significant classification of auditory conditions in superficial layers, there is a general trend towards higher accuracies in the deeper ROIs. This trend is evident when averaging across subjects (figure 4-3 C) with the three deepest ROIs significant (depth 6: 42.4%, p<0.0001; depth 5: 39.6%, p<0.0001; depth 4: 36.8%, p=0.0033). In the average auditory plot the most superficial depths are also above chance (depth 3: 37.5%,

p=0.0514; depth 2: 37.5%, p=0.0205; depth 1: 37.9%, p<0.0001), however visual inspection of the individual profiles suggest that this effect is driven mostly by a single subject. This particular subject also has significant results in deep layers.
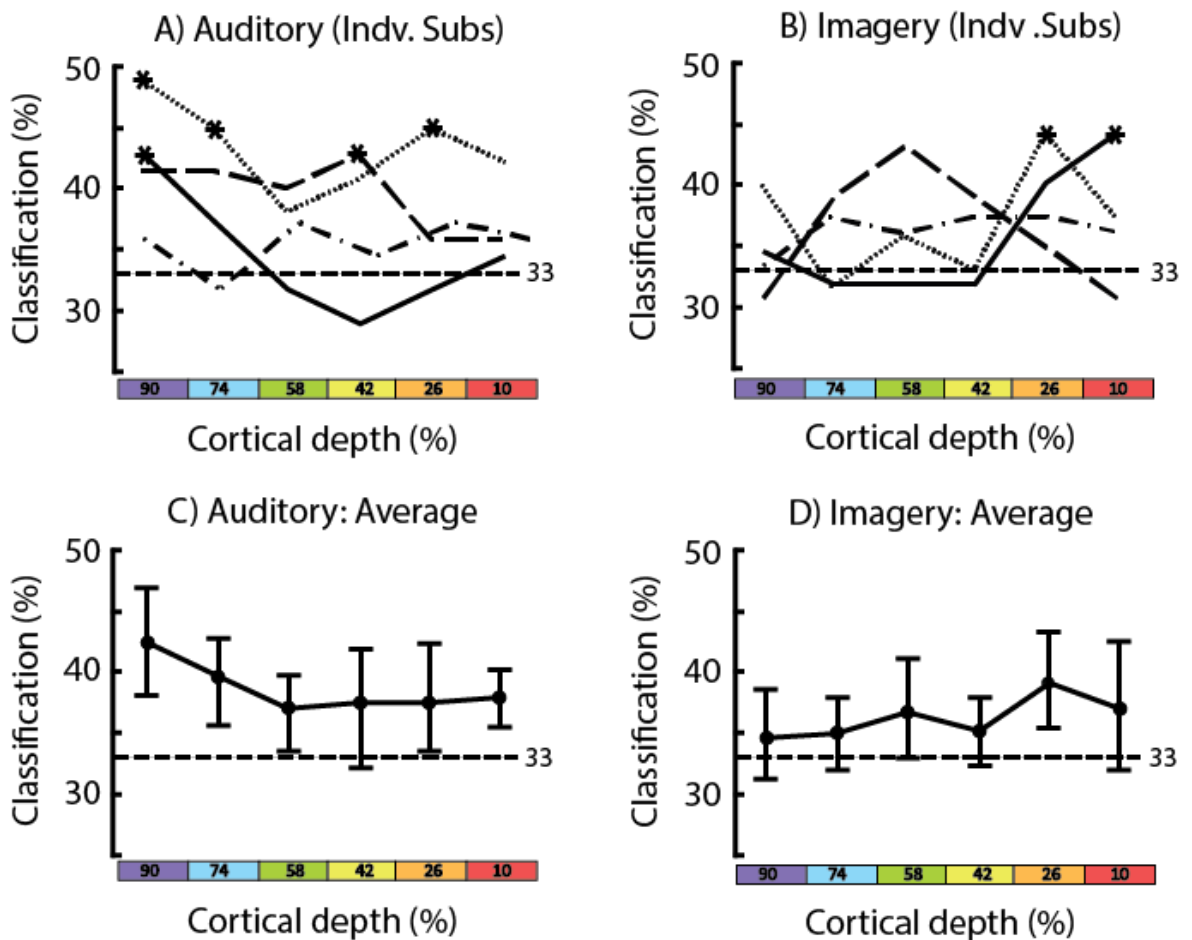
Figure 4-3 SVM decoding accuracies in V1 depth ROIs. **A.** Individual profiles for auditory conditions. **B.** Individual profiles for mental imagery conditions. **C.** Group average profile for auditory conditions. **D.** Group average profile for mental imagery conditions. Asterisks represent significant permutation tests. Group data was bootstrapped (with replacement n=10'000) to produce 95% confidence intervals.

In the case of imagery information, significant effects were found only in superficial depths of two subjects (figure 4-3 B) and these subjects clearly drive the effect in the average plot (figure 4-3 D) at depth 2. A different subject drives an effect in depth 4 (depth 6: 34.7%, p=0.2029; depth 5: 35.1%, p=0.0634; depth 4: 34.8%, p=0.0391; depth 3: 35.4%, p=0.0590; depth 2: 39.2%, p<0.0001; depth 1: 37.2%, p=0.0350).

4.2.4 Discussion

4.2.4.1 Feedback to Foveal and Peripheral V1

We recorded four subjects with high resolution fMRI at 7T using identical stimuli, task and procedures as in Vetter et al. (2014). We did not replicate either the far peripheral advantage for classification of auditory scene information or the foveal advantage for decoding mental imagery content. Instead, we found that we could classify auditory information in nearly every ROI. Classifying mental imagery was possible far peripheral cortex for V1, V2 and V3 and for peripheral V1 and foveal V2. There are several potential reasons for the discrepancy between our results and Vetter et al. (2014). First, Vetter et al. (2014) used many more subjects than we did in the current study and that it is possible that with a comparable number of subjects we would find a similar pattern across eccentricity. Another possibility is that the higher functional resolution was more sensitive to informative patterns at a smaller spatial scale.

There is some evidence from diffusion tensor imaging studies (Beer, Plank, & Greenlee, 2011; Beer, Plank, Meyer, & Greenlee, 2013) that connections from auditory cortices do reach all the way to foveal cortex (at tip of the occipital pole). In fact, Beer et al. (2011), found that connections to occipital pole and anterior calcarine sulcus were comparable in magnitude. Given that the existence of auditory feedback connections to peripheral visual cortex were not recognised for many years (Clavagnier et al., 2004), it is not unthinkable that there are still feedback connections to foveal cortex waiting to be discovered. Indeed, Falchier et al. (2002) did find a small number of foveal connections. It could be that 7T fMRI is sensitive enough to detect information delivered by these weaker connections whereas 3T fMRI is not.

4.2.4.2 Feedback to infra- and supra- granular layers

The higher functional resolution of 7T fMRI allows us to probe information across cortical depths. While the effective resolution is not high enough to distinguish signals emanating solely from one of the six anatomical layers (due to several factors that unavoidably lead to spatial smoothing of the data), we can nonetheless probe supra- and infra-granular cortical depths. This is important

because feedforward and feedback connections terminate with distinct laminar profiles relative to the granular cortical layer (Markov et al., 2013; Markov et al., 2014). Our data show a trend towards auditory information in the infra-granular depths and mental imagery content in supra-granular depths. The fact that these trend are opposite for auditory and imagery information (particularly when comparing the two subjects with solid and fine dashed lines in figure 4-3 A and B) is interesting because this could point to different feedback processes subserving these functions.

An extensive tracing study (Markov et al., 2013; Markov et al., 2014) using 26 monkeys to examine the connectivity among many high and low level areas in the visual cortical hierarchy found that feedback connections from distant cortical areas project to superficial layer 1 whereas adjacent and nearby cortical areas have feedback connections which arrive into both supra- and infra- granular layers. Further, in the supra-granular layer both feedforward and feedback connections are topographically organised with a similar degree of precision, whereas in infra-granular layers there is much less precision (i.e. connections originating nearby each other often terminate in very different locations). Why this should be the case is not known, but it maybe be that the degree of topographic precision involved in a recurrent processing operation is dependent on the perceptual task. If the trend for auditory information in the infra-granular layers of V1 is true, then this would imply that the information arrived into V1 not directly from a long range auditory cortical source, but instead arrived from a nearby visual area that was the last in a chain of feedback linking a multisensory higher area (such as STP) to V1 via several nearby visual areas. It would also imply that there is less topographical precision in the auditory feedback than of the imagery feedback. This prediction may not be testable without invasive physiological tests.

However, it must be kept in mind that the scheme suggested by Markov's data is based solely on data from the visual system, and may miss some features of the typical laminar connectivity profile between different sensory modalities. There is also the fact that even though inter-areal connections have such identifiable laminar profiles, there are many intra-areal connections between the different layers which serve to integrate feedback and feedforward signals (Andre M Bastos

et al., 2012) and as such characteristic patterns of activity in one layer could presumably bring about a related pattern in another layer (Iurilli et al., 2012). Therefore we must be cautious when interpreting laminar BOLD signals (and indirect measure of neural activity) as being due to the connectivity profile of feedforward and feedback connections among distinct areas.

4.2.4.3 Conclusions

Feedback of auditory information to early visual cortex was detected decades ago (Morrell, 1972) but its functional has remained unclear. Here we demonstrate that laminar investigation of feedback inputs to visual cortex is possible with 7T fMRI. Depth dependent classification profiles show that middle depths do not carry auditory information in line with a non-feedforward signal. Moreover, mental imagery showed a different laminar information profile than did auditory stimulation, suggesting that our auditory results were not due to subjects engaging in mental imagery. The suggestion that auditory feedback aids gaze/attentional orientating (Morrell, 1972) do not explain our finding of abstract auditory scene information. It could be that auditory feedback plays a more active role in visual cortex to disambiguate noisy visual inputs or perhaps to help synchronise information processing in parallel sensory systems.

4.3 Chapter Discussion

Working with ultra-high filed fMRI data is extremely challenging, particularly since common preprocessing and analysis conventions used with 3T fMRI may not be optimal for use with 7T data (Kashyap, Ivanov, Havlicek, Poser, & Uludağ, 2017; Self, van Kerkoerle, Goebel, & Roelfsema, 2017). Solutions and conventions are changing rapidly and for this reason the data presented in this chapter are not fully analysed. Some issues that become more problematic or more important to address at 7T are: correcting geometric distortions of the functional data due to magnetic susceptibility gradients, achieving good alignment of the functional to the anatomical data and producing a high quality segmentation of the anatomical MRI data. These challenges are all essential to overcome in order to conclude that the laminar differences we observe in classification accuracy are really due to functional differences in depth-dependent processing. Even a small misalignment

of a millimetre or so could shift our ROIs such that deep and superficial depths are actually sampling mid-depth grey matter. Another set of issues relate to smoothing of the data (Gardumi et al., 2016) which reduces the effective resolution and can occur due to: subject motion, preprocessing choices such as interpolation of the data into non-native spaces and analytical choices such as how best to model BOLD activity in different cortical depths (Lawrence, Formisano, Muckli, & de Lange, 2017; Uludağ & Blinder, 2017) and how best to sample voxels which do not respect the curvature of the grey matter sheet.

Given all of these complicating factors, it is important to realise that the often quoted functional resolution of $0.8mm^3$ is somewhat reduced by the aforementioned issues. Therefore we cannot and do not claim to be recording signals emanating solely from particular anatomical layer. However we can safely assume functional voxels and various cortical depths to represent a general bias to the processing carried out in the underlying anatomical layers. Moreover, new software tools are being developed to better deal with issues encountered in ultra-high field fMRI data. For instance, mesh boundary-based alignment methods might produce better alignment between the anatomy and functional data. Additionally, projecting the defined ROIs into the native functional space rather than unnecessarily interpolating the functional data into non-native spaces will avoid inherent smoothing of the functional data. These new tools should allow us to get a clearer and perhaps more consistent picture of the differences in the laminar profile of auditory and imagery information.

It is also clear from the variance among the laminar profiles of the individual subjects that more subjects are needed to clarify the effects we see in the current data. Feedback activity in infragranular layers has been reported in ultra-high field fMRI before by Kok et al. (2016) using a visual illusion paradigm. It is worth noting that the individual laminar profiles of individual subjects (of which there were ten) were largely inconsistent and the reported profile is only apparent in the average (figure S3 in Kok et al., 2016). The layer segmentations in that study relied exclusively on automated reconstruction algorithms, which may go some way in explaining the between-subject variance.

However, in the present study the fact that the same depth ROIs were used to classify the auditory and imagery conditions, and that auditory and imagery runs

were interleaved during the recording sessions mean that any differences in the laminar profile of classification accuracy between auditory and imagery conditions cannot be due to some confound involving due to differential alignment of the auditory and imagery runs to the anatomy or due to signal related differences across cortical depths.

In summary, we demonstrate trends of auditory and imagery representations in V1 deep and superficial layers, respectively. These results are promising, and with the development and optimisation of ultra-high field fMRI methodology and analysis, we may achieve a better understanding of this data.

5 General Discussion

The anatomical segregation between feedforward and feedback signals mean that they almost certainly perform different functional roles (Markov et al., 2014). Therefore, studying and understanding feedback in addition to feedforward processing is vital for understanding general brain function. In this thesis I have described three lines of investigation that focus on where feedback signals project to in early visual cortex and to what extent the information is suited to the cortical project site. This focus was broadly motivated by conspicuous biases in large scale retinotopy (chapter 2), known spatial frequency processing capabilities (chapter 3) and in anatomical connectivity profiles (chapter 4). We used a range of manipulations to explore this theme. In particular we varied task, the presence or absence of concurrent relevant feedforward information and the stimulus input modality. We found several cases in which foveal and peripheral early visual cortex received different feedback information.

In chapter 2 we saw that feedback of object information is projected to foveal V1 cortex – possibly with increased detail during an object identification task whilst background scene information is projected diffusely to fovea land peripheral V1 and can be disrupted by a sufficiently demanding object discrimination task. In chapter 3 we saw an indication that scene category perception was fed-back to far peripheral cortex whilst foveal cortex more faithfully represents the feedforward input regardless of perception. Taken together, these results could be taken to support the idea that retinotopic eccentricity biases throughout the visual hierarchy – the foveal and peripheral bias of higher object and scene areas, respectively – provide an organisational scheme for somewhat segregated cortical feedback of object and scene information. In chapter 4 we demonstrate that laminar investigation of feedback inputs to visual cortex is possible with 7T fMRI and find that infra- and supra-granular layers seem to receive auditory and mental imagery feedback, respectively. The functional significance of this observation is unclear but warrants further investigation.

Future work should address a number of limitations with the experiments in chapter 2. Drawing conclusions about feedback in this paradigm depends critically on the assumption that no differential feedforward information stimulated the

subsection of V1 being analysed. Even if a very small minority of voxels receive such stimulation, MVPA methods may capitalise on this. In fact, this could well mimic our observed results: slightly above chance classification (based on a limited feedforward signal). It is hard to know how successful we were in controlling for this possibility and with only three experiments, each with relatively small numbers of subjects, this poses a concern. However, the work presented represents only a subset of the total work done by our lab using this paradigm. In addition to the fact that the paradigm has been replicated a number of times within the lab, the variety of stimuli, analytical and data recording manipulations across all the experiments serve to increase confidence in the general conclusion. While each individual experiment within the lab would not be convincing on its own, taken together they form a coherent picture of feedback influences. However, replications within a lab likely contain a fair amount of overlap in the preprocessing and analytical procedures (at least this is the case here) and so possibly could be 'replicating' systematic biases and errors as well. That is why inter-lab replication is so important. On this front we can point to the fMRI results of Williams et al. (2008) in which object information was decoded in the fovea but not in the periphery and upon which we based the design of our experiments.

One of the major outstanding issues is that of eye movements – any characteristic eye movement associated with a condition might create outright an informative signal for the classifier, or might allow feedforward stimulation to flow into the 'occluded' ROI. While we did record eye movements in most of the subjects, we have not yet analysed the data. Following the example of published work from other labs, my lab colleagues have generally plotted the average gaze position and shown it to be tightly concentrated on the central fixation point, discarding any volumes in which the gaze deviated substantially. However, this may not be sufficient – if micro-saccades occurred in a consistently different way for different conditions, the fixation point will be 'jittered on the retina' inducing an informative pattern of (feedforward) stimulation in the cortex. A good test to pass for our data would be an attempt to use an SVM on the eye-tracking data to classify the condition directly – if this is possible then it would demonstrate that the subject's eyes did indeed move differently between conditions. The results from eye-tracking SVMs and BOLD SVMs could be correlated across subjects to test the

notion that eye movements are responsible for our fMRI data and associated results.

There are some issues common to all the experiments. The most fundamental is probably the low number of subjects. Increasing this would have given us the statistical power to detect or reject effects conclusively and thus make stronger and unambiguous inferences that we could be confident would generalise to the population at large. Almost all of the analysis presented in this thesis is fairly shallow – simply bootstrapping the classification accuracies across subjects, and in some cases does not take into account the number of comparisons being made (which tends to increase the false positive error rate). Another statistical error is not taking into account the random effects associated with our subjects (this is a form of overfitting), which again weakens the generalisability of our results to the general population. Linear mixed effects models have gained popularity recently and represent a powerful and elegant approach to modelling complex data sets. They can handle subjects as random effects to make generalisation possible – even when including other random effects in the model (Barr, Levy, Scheepers, & Tily, 2013).

In general, however, the small sample sizes we have worked with here may be better suited to some form of Bayesian analysis (McNeish, 2016; Van De Schoot, Broere, Perryck, Zondervan-Zwijnenburg, & Van Loey, 2015), which takes advantage of prior information (what is already 'known') and updates our knowledge using the collected data. While growing in popularity, Bayesian approaches are still not common practice in the field – although this is not a good argument against their use. However, they do require a new set of expertise to apply correctly and misunderstanding can lead to severe problems (McNeish, 2016). A central issue when trying to apply Bayesian analysis to small sample sizes centres around specifying appropriate priors – which can entail judgements that are not easily agreed upon (Van De Schoot et al., 2015). Bad judgements in such cases can actually give more unreliable results than conventional frequentist analyses and it is not the case (although it is commonly assumed) that diffuse or uninformative priors are a safe option which at worse will be roughly equivalent to frequentist statistics (McNeish, 2016). None of this argues against using

Bayesian methods, it merely points out that a substantial learning investment or help from an expert is required to carry it out successfully.

Another question relates to what we are actually measuring in these experiments – that is, what information does successful classification by an SVM reveal to us? Anderson and Oates (2010) discuss this issue and lay bare the complexities and subtleties of the answer – which may be summarised as 'not what many people in the fMRI community infer from them'. More specifically, it does not give a straight forward localisation of which voxels contain the most information within a considered ROI (e.g. by simply looking at the voxels with the highest weights etc.). Our experiments are safe from this criticism, since all we claim is that there is some information residing, in some form, somewhere in an ROI – it is the nature of the ROI itself which is interesting in chapters 2 and 4: V1, but in the absence of informative feedforward input. From there we conclude that the existence of such information is revealing what the brain has to work with (i.e. feedback signals) for its subsequent computations. Here though, we rely on another assumption that may not be valid – just because an SVM (or other MVPA method) can extract information from a BOLD signal does not mean that the brain is or even could be making use of it. Although this point could also be made about univariate analyses (a difference in average activation doesn't necessarily imply that the brain is making use of the implied neural difference). It would be beneficial to relate the MVPA results to behavioural measures or to other brain areas to build a stronger case that the degree to which information is available to the experimenter is actually of consequence to the functioning of the brain itself (de-Wit, Alexander, Ekroll, & Wagemans, 2016). For our research, we do touch upon this idea (in chapter 2 experiment 3) where we analyse the representational similarity between ROIs. Although, as mentioned above, we could have given more careful thought about the task to be done by the subject (beyond simply giving them a task to keep them attending the stimuli).

Another consideration, once MVPA is going to be pursued, is which method to use to quantify the dissimilarity between two given patterns – SVMs and other classifiers are popular, but so are various 'distance measures' which don't take the additional step of actually classifying the data (examples are Euclidian distance, Pearsons correlation etc.). Also important decisions are what steps to

include in computing the chosen dissimilarity metric. Walther et al. (2016) conducted simulations to show that cross-validated Mahalanobis distance with prior multivariate noise normalisation of the data gives the most reliable results. This metric has the added advantage that it is more interpretable – it has a true zero point (so a particular dissimilarity measurement can be meaningfully said to be twice as large as another, whereas it is not straightforward to say that an SVM which is 20% above chance represents a dissimilarity twice as large as an SVM 10% above chance – the underlying dissimilarities may be more or less than this). This metric also satisfies parametric assumptions and is therefore are more amenable to conventional inferential statistics.

By addressing these limitations and implementing the improvements to the analysis, future work can build on the results presented in this thesis in an attempt to clarify the purpose of feedback to early visual cortex. To achieve this, we may need focus more on designing appropriate tasks that engage a variety of complex perceptual processes and go beyond simple discrimination and recognition. Different tasks stress different aspects of information processing. For example, a fine visual discrimination task may require careful processing of high spatial frequency information, whereas a rapid visual categorisation task may be best be performed using low spatial frequency information. V1 must be the starting point of processing for both these tasks and it makes sense that to maximise task performance, processing at the earliest levels should be adaptable. As such, feedback might usefully act to request and guide computations that are useful to the task of perception about the data, perhaps requesting early visual cortices to dedicate more processing resources onto relevant aspects of the data which remain unresolved in the higher areas.

One popular explanation for the functional purpose of feedback is predictive coding, which posits that each area in the cortical hierarchy attempts to create an internal model to explain the feedforward data sent to it by earlier areas (Andre M Bastos et al., 2012; Friston, 2010). Predictions of this feedforward data are then sent through feedback to earlier areas. Only feedforward data not successfully predicted by the model is sent on to the higher area, as this data shows where the model is incorrect. Errors in prediction of feedforward data that is highly precise and thus more reliable are given more weight. On receiving these

prediction errors, the higher area attempts to refine its model. There is also a suggestion that the brain actively tries to sample new input data as means of testing a wide range of model predictions ('active hypothesis testing') to which results in better models. Thus on this view the brain is seen as a prediction machine which actively continuously tries to minimise prediction errors (Clark, 2013).

This is widely thought to be an elegant idea which offers a fundamental principle of possible cortical function. While it does incorporate most major sensory tasks, and while our data could be viewed within such a framework, there is still a gulf between the theoretical components of the predictive coding framework and what can be directly measured by experiment. This necessitates a long chain of reasoning to derive concrete predictions about the outcome of a particular experiment. Consequently, it is often hard to see experimental data as strongly supporting or contradicting the original idea. In a commentary on a piece by (Clark, 2013), Philips points out that a simple but very real and major challenge for any 'unifying theory' of cortical function is the sheer diversity of brains and cognition and concludes that although possible in principle, in practice dealing with this challenge is largely a hope for the future.

We can also question whether the brain should be expected to operate according to an elegant fundamental principle and thus whether elegance in theory should be highly valued in a theory of cortical function. The brain is an evolved organ and while evolution often produces effective and efficient solutions, it must do so through a series of gradual adaptions to previous designs each of which functions. This process often results in a well-functioning but inherently patchwork system. This consideration does not argue against predictive coding per se, but does highlight the need to consider that cortical function may operate through an efficient 'bag of tricks' (Dennett, 2017) rather than a single fundamental principle.

While it is difficult to produce data strongly for or against predictive coding, its all-encompassing nature makes it a useful framework within which to compare other models. It also highlights the role of feedback within cortical function, providing an important counterbalance to the historical domination of much

theoretical and empirical work by the so called feed-forward hierarchical approach.

References

Al-Aidroos, N., Said, C. P., & Turk-Browne, N. B. (2012). Top-down attention switches coupling between low-level and high-level areas of human visual cortex. *Proceedings of the National Academy of Sciences, 109*(36), 14675-14680.

Anderson, M., & Oates, T. (2010). *A critique of multi-voxel pattern analysis.* Paper presented at the Proceedings of the Annual Meeting of the Cognitive Science Society.

Andersson, J. L., Skare, S., & Ashburner, J. (2003). How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *Neuroimage, 20*(2), 870-888.

Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in brain research, 154*, 93-120.

Aubert, H., & Foerster, C. (1857). Beitrage zur Kenntnisse der indirecten Sehens [Translation: Contributions of knowledge to indirect vision]. *Graefes Archiv fur Ophthalmologie, 3*, 1-37.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language, 68*(3), 255-278.

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., & Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron, 76*(4), 695-711.

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J.-M., Oostenveld, R., Dowdall, J. R., . . . Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron, 85*(2), 390-401.

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, 213*(2-3), 299.

Beer, A. L., Plank, T., Meyer, G., & Greenlee, M. W. (2013). Combined diffusion-weighted and functional magnetic resonance imaging reveals a temporal-occipital network involved in auditory-visual object processing. *Frontiers in integrative neuroscience, 7.*

Budd, J. M. (1998). Extrastriate feedback to primary visual cortex in primates: a quantitative analysis of connectivity. *Proceedings of the Royal Society of London B: Biological Sciences, 265*(1400), 1037-1044.

Burton, H. (2003). Visual cortex activity in early and late blind people. *Journal of Neuroscience, 23*(10), 4005-4011.

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., . . . Rust, N. C. (2005). Do we know what the early visual system does? *Journal of Neuroscience, 25*(46), 10577-10597.

Chambers, C. D., Allen, C. P. G., Maizey, L., & Williams, M. A. (2013). Is delayed foveal feedback critical for extra-foveal perception? *Cortex, 49*(1), 327-335. doi:10.1016/j.cortex.2012.03.007

Chen, M., Yan, Y., Gong, X., Gilbert, C. D., Liang, H., & Li, W. (2014). Incremental integration of global contours through interplay between visual cortical areas. *Neuron, 82*(3), 682-694.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181-204.

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), 117-126.

Cowey, A., & Rolls, E. (1974). Human cortical magnification factor and its relation to visual acuity. *Experimental Brain Research, 21*(5), 447-454.

Cukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience, 16*(6), 763.

Daniel, P., & Whitteridge, D. (1961). The representation of the visual field on the cerebral cortex in monkeys. *The Journal of physiology, 159*(2), 203-221.

de-Wit, L., Alexander, D., Ekroll, V., & Wagemans, J. (2016). Is neuroimaging measuring information in the brain? *Psychonomic Bulletin & Review, 23*(5), 1415-1428.

De Valois, R. L., Albrecht, D. G., & Thorell, L. G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. *Vision research, 22*(5), 545-559.

Dennett, D. C. (2017). *Consciousness explained*: Little, Brown.

Drasdo, N. (1977). The neural representation of visual space. *Nature, 266*(5602), 554.

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *Neuroimage, 39*(2), 647-660.

Duncan, R. O., & Boynton, G. M. (2003). Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron, 38*(4), 659-671.

Edwards, G., Vetter, P., McGruer, F., Petro, L. S., & Muckli, L. (2017). Predictive feedback to V1 dynamically updates with sensory input. *bioRxiv*, 180539.

Etzel, J. A., & Braver, T. S. (2013). MVPA Permutation Schemes.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience, 22*(13), 5749-5759.

Fan, X., Wang, L., Shao, H., Kersten, D., & He, S. (2016). Temporally flexible feedback signal to foveal cortex for peripheral object recognition. *Proceedings of the National Academy of Sciences, 113*(41), 11627-11632.

Fang, F., Boyaci, H., Kersten, D., & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. *Current Biology, 18*(21), 1707-1712.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991), 1*(1), 1-47.

Fishman, M. C., & Michael, C. R. (1973). Integration of auditory information in the cat's visual cortex. *Vision research, 13*(8), 1415-1419.

Fracasso, A., Petridou, N., & Dumoulin, S. O. (2016). Systematic variation of population receptive field properties across cortical depth in human visual cortex. *Neuroimage, 139*, 427-438.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Gaglianese, A., Vansteensel, M. J., Harvey, B. M., Dumoulin, S. O., Petridou, N., & Ramsey, N. F. (2017). Correspondence between fMRI and electrophysiology during visual motion processing in human MT+. *Neuroimage*.

Gardumi, A., Ivanov, D., Hausfeld, L., Valente, G., Formisano, E., & Uludağ, K. (2016). The effect of spatial resolution on decoding accuracy in fMRI multivariate pattern analysis. *Neuroimage, 132*, 32-42.

Goense, J. B., & Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology, 18*(9), 631-640.

Grill-Spector, K. (2003). The neural basis of object perception. *Current opinion in neurobiology, 13*(2), 159-166.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision research, 41*(10), 1409-1422.

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

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature reviews. Neuroscience, 15*(8), 536.

Harel, A., Kravitz, D. J., & Baker, C. I. (2013). Deconstructing visual scenes in cortex: gradients of object and spatial layout information. *Cerebral Cortex, 23*(4), 947-957.

Harvey, B. M., Vansteensel, M. J., Ferrier, C. H., Petridou, N., Zuiderbaan, W., Aarnoutse, E. J., . . . Leijten, F. S. (2013). Frequency specific spatial interactions in human electrocorticography: V1 alpha oscillations reflect surround suppression. *Neuroimage, 65*, 424-432.

Hasson, U., Harel, M., Levy, I., & Malach, R. (2003). Large-scale mirror-symmetry organization of human occipito-temporal object areas. *Neuron, 37*(6), 1027-1041.

Hasson, U., Levy, I., Behrmann, M., Hendler, T., & Malach, R. (2002). Eccentricity bias as an organizing principle for human high-order object areas. *Neuron, 34*(3), 479-490.

Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences, 7*(11), 498-504.

Henriksson, L., Khaligh-Razavi, S.-M., Kay, K., & Kriegeskorte, N. (2015). Visual representations are dominated by intrinsic fluctuations correlated between areas. *Neuroimage, 114*, 275-286.

Hermes, D., Miller, K. J., Vansteensel, M. J., Aarnoutse, E. J., Leijten, F. S., & Ramsey, N. F. (2012). Neurophysiologic correlates of fMRI in human motor cortex. *Human brain mapping, 33*(7), 1689-1699.

Hilz, R., & Cavonius, C. (1974). Functional organization of the peripheral retina: Sensitivity to periodic stimuli. *Vision research, 14*(12), 1333-1337.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791-804.

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), 814-828.

Jacques, C., Witthoft, N., Weiner, K. S., Foster, B. L., Rangarajan, V., Hermes, D., . . . Grill-Spector, K. (2016). Corresponding ECoG and fMRI category-selective signals in human ventral temporal cortex. *Neuropsychologia, 83*, 14-28.

Johnson, C. A., & Scobey, R. P. (1980). Foveal and peripheral displacement thresholds as a function of stimulus luminance, line length and duration of movement. *Vision research, 20*(8), 709-715.

Jüttner, M., & Rentschler, I. (1996). Reduced perceptual dimensionality in extrafoveal vision. *Vision research, 36*(7), 1007-1022.

Kashyap, S., Ivanov, D., Havlicek, M., Poser, B. A., & Uludağ, K. (2017). Impact of acquisition and analysis strategies on cortical depth-dependent fMRI. *Neuroimage*.

Kauffmann, L., Ramanoël, S., Guyader, N., Chauvin, A., & Peyrin, C. (2015). Spatial frequency processing in scene-selective cortical regions. *Neuroimage, 112*, 86-95.

Kayser, C., Körding, K. P., & König, P. (2004). Processing of complex stimuli and natural scenes in the visual cortex. *Current opinion in neurobiology, 14*(4), 468-473.

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current Biology*.

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), 3863-3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience, 2*, 4.

Larkum, M. (2013). A cellular mechanism for cortical associations: an organizing principle for the cerebral cortex. *Trends in neurosciences, 36*(3), 141-151.

Larson, A. M., & Loschky, L. C. (2009). The contributions of central versus peripheral vision to scene gist recognition. *J Vis, 9*(10), 6-6.

Lawrence, S. J., Formisano, E., Muckli, L., & de Lange, F. P. (2017). Laminar fMRI: applications for cognitive neuroscience. *Neuroimage*.

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.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision research, 48*(5), 635-654.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. *Nature Neuroscience, 4*(5), 533-539.

Levy, I., Hasson, U., Harel, M., & Malach, R. (2004). Functional analysis of the periphery effect in human building related areas. *Human brain mapping, 22*(1), 15-26.

Logothetis, N. K. (2010). Bold claims for optogenetics. *Nature, 468*(7323), E3-E4.

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.

Ludvigh, E. (1941). Effect of reduced contrast on visual acuity as measured with Snellen test letters. *Archives of Ophthalmology, 25*(3), 469-474.

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. *Journal of Neuroscience, 32*(4), 1395-1407.

Malach, R., Levy, I., & Hasson, U. (2002). The topography of high-order human object areas. *Trends in cognitive sciences, 6*(4), 176-184.

Markov, N. T., Ercsey-Ravasz, M., Lamy, C., Gomes, A. R. R., Magrou, L., Misery, P., . . . Toroczkai, Z. (2013). The role of long-range connections on the specificity of the macaque interareal cortical network. *Proceedings of the National Academy of Sciences, 110*(13), 5187-5192.

Markov, N. T., Vezoli, J., Chameau, P., Falchier, A., Quilodran, R., Huissoud, C., . . . Ullman, S. (2014). Anatomy of hierarchy: feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology, 522*(1), 225-259.

McNeish, D. (2016). On using Bayesian methods to address small sample problems. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(5), 750-773.

Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J.-M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron, 89*(2), 384-397.

Morrell, F. (1972). Visual system's view of acoustic space. *Nature, 238*(5358), 44-46.

Muckli, L. (2010). What are we missing here? Brain imaging evidence for higher cognitive functions in primary visual cortex V1. *International Journal of Imaging Systems and Technology, 20*(2), 131-139.

Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., . . . Yacoub, E. (2015). Contextual feedback to superficial layers of V1. *Current Biology, 25*(20), 2690-2695.

Muckli, L., & Petro, L. S. (2013). Network interactions: Non-geniculate input to V1. *Current opinion in neurobiology, 23*(2), 195-201.

Müller, J. R., Metha, A. B., Krauskopf, J., & Lennie, P. (1999). Rapid adaptation in visual cortex to the structure of images. *Science, 285*(5432), 1405-1408.

Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience, 9*(3), 429.

Musel, B., Bordier, C., Dojat, M., Pichat, C., Chokron, S., Le Bas, J.-F., & Peyrin, C. (2013). Retinotopic and lateralized processing of spatial frequencies in human visual cortex during scene categorization. *Journal of cognitive neuroscience, 25*(8), 1315-1331.

Musel, B., Kauffmann, L., Ramanoël, S., Giavarini, C., Guyader, N., Chauvin, A., & Peyrin, C. (2014). Coarse-to-fine categorization of visual scenes in scene-selective cortex. *Journal of cognitive neuroscience, 26*(10), 2287-2297.

Nassi, J. J., Lomber, S. G., & Born, R. T. (2013). Corticocortical feedback contributes to surround suppression in V1 of the alert primate. *Journal of Neuroscience, 33*(19), 8504-8517.

Nuthmann, A., & Henderson, J. M. (2010). Object-based attentional selection in scene viewing. *J Vis, 10*(8), 20-20.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive psychology, 34*, 72-107.

Oliva, A., Torralba, A., & Schyns, P. G. (2006). *Hybrid images.* Paper presented at the ACM Transactions on Graphics (TOG).

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding V1? *Neural computation, 17*(8), 1665-1699.

Özgen, E., Sowden, P., Schyns, P., & Daoutis, C. (2005). Top-down attentional modulation of spatial frequency processing in scene perception. *Visual Cognition, 12*(6), 925-937.

Perkel, D. J., Bullier, J., & Kennedy, H. (1986). Topography of the afferent connectivity of area 17 in the macaque monkey: A double-labelling study. *Journal of Comparative Neurology, 253*(3), 374-402.

Petro, L., Paton, A., & Muckli, L. (2017). Contextual modulation of primary visual cortex by auditory signals. *Phil. Trans. R. Soc. B, 372*(1714), 20160104.

Petro, L. S., Vizioli, L., & Muckli, L. (2014). Contributions of cortical feedback to sensory processing in primary visual cortex. *Frontiers in psychology, 5*.

Peyrin, C., Baciu, M., Segebarth, C., & Marendaz, C. (2004). Cerebral regions and hemispheric specialization for processing spatial frequencies during natural scene recognition. An event-related fMRI study. *Neuroimage, 23*(2), 698-707.

Peyrin, C., Chokron, S., Guyader, N., Gout, O., Moret, J., & Marendaz, C. (2006). Neural correlates of spatial frequency processing: A neuropsychological approach. *Brain Research, 1073*, 1-10.

Peyrin, C., Schwartz, S., Seghier, M., Michel, C., Landis, T., & Vuilleumier, P. (2005). Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *Neuroimage, 28*(2), 464-473.

Poort, J., Raudies, F., Wannig, A., Lamme, V. A., Neumann, H., & Roelfsema, P. R. (2012). The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron, 75*(1), 143-156.

Poort, J., Self, M. W., van Vugt, B., Malkki, H., & Roelfsema, P. R. (2016). Texture segregation causes early figure enhancement and later ground suppression in areas V1 and V4 of visual cortex. *Cerebral Cortex, 26*(10), 3964-3976.

Rajimehr, R., Devaney, K. J., Bilenko, N. Y., Young, J. C., & Tootell, R. B. (2011). The "parahippocampal place area" responds preferentially to high spatial frequencies in humans and monkeys. *PLoS biology, 9*(4), e1000608.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2*(1).

Rockland, K. S. (2017). Chapter 1 - Anatomy of the Cerebral Cortex A2 - Cechetto, David F. In N. Weishaupt (Ed.), *The Cerebral Cortex in Neurodegenerative and Neuropsychiatric Disorders* (pp. 3-36). San Diego: Academic Press.

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology, 50*(1), 19-26.

Rovamo, J., & Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research, 37*(3), 495-510.

Rovamo, J., Virsu, V., & Näsänen, R. (1978). Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. *Nature, 271*(5640), 54.

Salin, P.-A., & Bullier, J. (1995). Corticocortical connections in the visual system: structure and function. *Physiological reviews, 75*(1), 107-155.

Scholte, H. S., Jolij, J., Fahrenfort, J. J., & Lamme, V. A. (2008). Feedforward and recurrent processing in scene segmentation: electroencephalography

and functional magnetic resonance imaging. *Journal of cognitive neuroscience, 20*(11), 2097-2109.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time-and spatial-scale-dependent scene recognition. *Psychological science, 5*(4), 195-200.

Self, M. W., van Kerkoerle, T., Goebel, R., & Roelfsema, P. R. (2017). Benchmarking laminar fMRI: Neuronal spiking and synaptic activity during top-down and bottom-up processing in the different layers of cortex. *Neuroimage*.

Siero, J. C., Hermes, D., Hoogduin, H., Luijten, P. R., Petridou, N., & Ramsey, N. F. (2013). BOLD consistently matches electrophysiology in human sensorimotor cortex at increasing movement rates: a combined 7T fMRI and ECoG study on neurovascular coupling. *Journal of Cerebral Blood Flow & Metabolism, 33*(9), 1448-1456.

Siero, J. C., Hermes, D., Hoogduin, H., Luijten, P. R., Ramsey, N. F., & Petridou, N. (2014). BOLD matches neuronal activity at the mm scale: a combined 7T fMRI and ECoG study in human sensorimotor cortex. *Neuroimage, 101*, 177-184.

Smith, F. W., & Muckli, L. (2010). Nonstimulated early visual areas carry information about surrounding context. *Proceedings of the National Academy of Sciences, 107*(46), 20099-20103.

Sowden, P. T., Özgen, E., Schyns, P. G., & Daoutis, C. (2003). Expectancy effects on spatial frequency processing. *Vision research, 43*(26), 2759-2772.

Sowden, P. T., & Schyns, P. G. (2006). Channel surfing in the visual brain. *Trends in cognitive sciences, 10*(12), 538-545.

Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & psychophysics, 49*(6), 495-508.

Strasburger, H., Rentschler, I., & Harvey Jr, L. O. (1994). Cortical magnification theory fails to predict visual recognition. *European Journal of Neuroscience, 6*(10), 1583-1588.

Tootell, R. B., Silverman, M., Hamilton, S., Switkes, E., & De Valois, R. (1988). Functional anatomy of macaque striate cortex. V. Spatial frequency. *Journal of Neuroscience, 8*(5), 1610-1624.

Uludağ, K., & Blinder, P. (2017). Linking brain vascular physiology to hemodynamic response in ultra-high field MRI. *Neuroimage*.

Van De Schoot, R., Broere, J. J., Perryck, K. H., Zondervan-Zwijnenburg, M., & Van Loey, N. E. (2015). Analyzing small data sets using Bayesian estimation: The case of posttraumatic stress symptoms following mechanical ventilation in burn survivors. *European Journal of Psychotraumatology, 6*(1), 25216.

Van Essen, D., Newsome, W., Maunsell, J., & Bixby, J. (1986). The projections from striate cortex (V1) to areas V2 and V3 in the macaque monkey: asymmetries, areal boundaries, and patchy connections. *Journal of Comparative Neurology, 244*(4), 451-480.

Vetter, P., Smith, F. W., & Muckli, L. (2014). Decoding sound and imagery content in early visual cortex. *Current Biology, 24*(11), 1256-1262.

Virsu, V., & Rovamo, J. (1979). Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research, 37*(3), 475-494.

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage, 137*, 188-200.

Wang, P., & Cottrell, G. (2016). Modeling the contribution of central versus peripheral vision in scene, object, and face recognition. *arXiv preprint arXiv:1604.07457.*

Weymouth, F. W. (1958). Visual sensory units and the minimal angle of resolution. *American journal of ophthalmology, 46*(1), 102-113.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior research methods, 42*(3), 671-684.

Williams, M. A., Baker, C. I., Op de Beeck, H. P., Shim, W. M., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience, 11*(12), 1439-1445. doi:10.1038/nn.2218

Wokke, M. E., Vandenbroucke, A. R., Scholte, H. S., & Lamme, V. A. (2013). Confuse your illusion: feedback to early visual cortex contributes to perceptual completion. *Psychological science, 24*(1), 63-71.

Xing, J., & Heeger, D. J. (2000). Center-surround interactions in foveal and peripheral vision. *Vision research, 40*(22), 3065-3072.

Zeidman, P., Mullally, S. L., Schwarzkopf, D. S., & Maguire, E. A. (2012). Exploring the parahippocampal cortex response to high and low spatial frequency spaces. *Neuroreport, 23*(8), 503.