The neural basis of binocular depth perception

by

Matthew Lindsay Patten

A thesis submitted to the University of Birmingham

for the degree of Doctor of Philosophy.

November, 2012

School of Psychology

College of Life and Environmental Sciences

University of Birmingham

Abstract

How does the human visual system convert two-dimensional projections from our eyes into a three-dimensional percept? One primary method is from binocular disparities, which result from having two horizontally separated eyes and are used to provide a powerful cue to depth in our environment. In this thesis, I use human fMRI to investigate the cortical signals associated with binocular disparity. I address several core issues, including the relationship between cortical activity and perception, the significance of the reference plane on depth configurations, and the topography of disparity signals on the cortical surface.

In measuring responses to coarse and fine disparities, researchers typically engage two respective tasks: a signal-in-noise and a feature difference task. In the first chapter, we decouple the disparity magnitude from the perceptual task and examine cortical responses to both of these tasks when using fine disparities. Further, we manipulated performance and identified visual areas whose activity varied in line with perceptual judgments. We reveal that responses in later dorsal regions VIPS and POIPS were closely related to perception for both tasks. In the second chapter, we used a similar manipulation to investigate cortical regions that have solved the correspondence problem and whose responses were consistent with the depth percept of the observer, and reveal that this takes place in V7 and VIPS.

The third chapter examines the importance of the reference in disparity calculations. We performed several classifications based on depths that were considered relative to fixation or relative to the surround. We found that early visual areas were most sensitive to disparity edges; dorsal visual areas used both the fixation plane and the surround in computing disparity whereas ventral visual areas processed disparity with reference to the surround.

In the fourth chapter, we attempt to identify a topographic organisation of binocular disparity in the visual cortex. We estimate the disparity preferences of each voxel in two distinct ways, and displayed these preferences on a flatmap of the cortical surface. Although

we did not observe a topographic map of disparity, we observed a cluster in intermediate dorsal regions (V3A, V3B/KO, V7) that consistently showed a bias towards crossed disparities of a larger magnitude.

Acknowledgements

I would like to thank Andrew Welchman for his guidance, training, support and good humour throughout the course of my Ph.D. I thank you for having patience with me and for always looking out for me.

I would like to thank Hiroshi Ban for his involvement in these projects and for saving me countless hours through help with programming, explaining concepts and discussing results. It didn't matter the question, you always had the answer. It didn't matter the time, you were always willing to help.

I would like to thank Zoe Kourtzi for guiding fMRI analyses and for helpful comments on numerous manuscripts, posters and presentations.

I would like to thank Satoshi Endo, Arthur Lugtigheid, Matthew Dexter, Alan Meeson, Aidan Murphy, Dicle Dövencioğlu, Christoph Böhme, Jiaxiang Zhang, Shuguang Kuai, Dorita Chang, Sheng Li and Steve Mayhew for technical assistance and/or scientific discussions that have developed my understanding and interest over a range of scientific concepts.

Thank you to the endless support from Jenni & Peter Cornegé, Jane Patten, Bruce & Candelle Patten, Michael & Jennifer Patten and Brad & Elyse Patten.

I would also like to thank my friends and colleagues for supporting me not only over these past few years, but also for the many more before that. I'm sorry that there's not enough space to mention you all by name.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3D	Three-dimensional.
(ph)AIP	(Putative human) anterior intraparietal area.
ANOVA	Analysis of variance.
BOLD	Blood-oxygenation-level-dependence.
CIP	Caudal intraparietal sulcus.
CSF	Cerebrospinal fluid.
EEG	Electroencephalography.
EPI	Echo-planar imaging.
fMRI	Functional magnetic resonance imaging.
fSNR	Functional signal-to-noise ratio.
GLM	General linear model.
hMT+/V5	The human homologue of the middle temporal region (also known as V5).
IT	Inferior temporal cortex.
LIP	Lateral intraparietal sulcus.
LO	The lateral occipital region.
MST	Medial superior temporal area.
MT	Medial temporal area.
MRI	Magnetic resonance imaging.
MVPA	Multi-voxel pattern analysis.
POIPS	The parieto-occipital region of the intraparietal sulcus.
PSC	Percent signal change.
RDS	Random dot stereogram.
RF	Radio frequency.
ROI	Region-of-interest.
SVM	Support vector machine.
TE	Echo time.
TEs	An area in the lower bank of the superior temporal sulcus.
TR	Repetition time.
V3B/KO	Kinetic occipital region (anatomically defined as V3B).
VIPS	The ventral part of the intraparietal sulcus.

1. General Introduction

1.1 Vision and depth perception

Of all the senses in the body, humans are largely dependent on vision. We use vision to build a representation of our surrounding environment, and use this representation to interact with and respond to the world around us. For example, through visual inspection we identify the transparent cylinder in front of us as a glass. We then use vision to direct our hand around the surface and grasp the object, bringing it towards us. Such a seemingly simple task requires a series of complex calculations, each of which is performed seemingly effortlessly and with meticulous precision by our visual system.

As already evident in the previous example, one of the fundamental requirements of our visual system is to calculate depth in our environment. An understanding of the three-dimensional (3D) layout of our surrounding is necessary in order to determine the spatial relationship between objects, others and ourselves. It allows us to navigate through an environment safely (e.g., by avoiding a tree branch; or grabbing onto a rock when climbing), facilitates fine motor actions (e.g., unlocking a door) and guides in the interaction with others (e.g., a handshake). However, the image formed from each eye is a two-dimensional projection of a three-dimensional world and by itself lacks a direct representation of the third-dimension. Our visual system has evolved to compute depth from the monocular cues that are present in the images of either one of the eyes. This can calculate depth indirectly through assumptions based on object size, shading, perspective, texture and occlusion, amongst others. However, we are able to perceive a much more vivid representation of depth from binocular cues, where we simultaneously inspect and combine the images from the two eyes.

1.2 Binocular disparity

To highlight the importance of binocular depth in humans, consider that many species have two eyes located on opposite sides of their head, greatly increasing the size of their visual field. However, primates (including humans) contain two eyes that are frontally placed and simultaneously inspect the same region of visual space. This trade-off substantially reduces our visual field but provides rich depth information through binocular disparities: the slight differences in the image of each eye caused by their horizontal offset (Howard & Rogers, 2002; Wheatstone, 1838). This directly relates to the location of objects in depth and is very precise in its representation of 3D space, with sensitivity reaching a few arcsec in ideal circumstances (i.e., a relative depth presented at fixation and viewed by an experienced observer) (McKee, 1983; McKee, Levi, & Bowne, 1990; Stevenson, Cormack, & Schor, 1989). Further, depth is perceived even when monocular information is uninformative to depth position, such as in random-dot stereograms (RDS; an array of randomly-positioned dots where a section of the dots contain a horizontal shift between the two eyes), revealing that binocular disparity alone is sufficient for stereopsis (Julesz, 1971). Aside from binocular disparities, stereopsis is also benefited from several other binocular cues including motion parallax, vergence eye-movements and motion in-depth.

We can think about the brain's representation of disparity signals with respect to two frames of reference. First, the absolute disparity of an object, computed through the horizontal difference in the images of the left and right retinae with respect to the foveae. Second, the relative disparity between two objects, calculated through the difference in their respective absolute disparities and independent of where the eyes are fixating. As illustrated in **Figure 1.1**, an observer that is fixating at point F has images which fall on corresponding points in the two eyes (the foveae). An object that lies at point P in the visual field subtends different

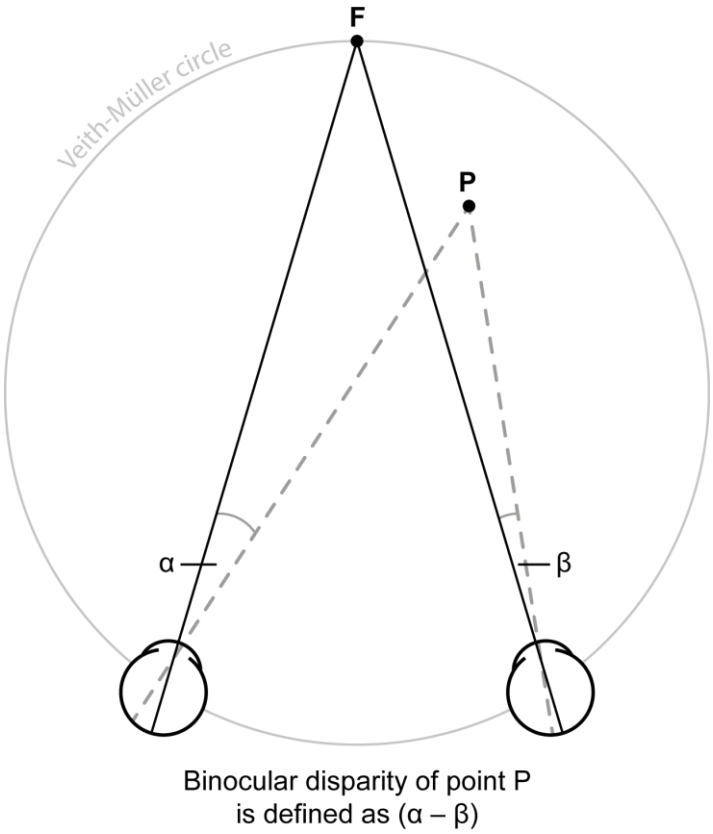


Figure 1.1: Binocular disparity. A top-down view of an observer who is fixating at point F with an object located at point P in the visual field. The projected images of P fall on different parts of the two retinae, and relative disparity is calculated by the difference in angles between two points (e.g., α – β). The large grey circle is the Veith-Müller circle and is the locus of points in space that have a binocular disparity of zero.

angles between the two eyes relative to this point (fixation), and the relative disparity is calculated through the difference between these angles (α – β). This is true regardless of where the eyes are fixating. As such, relative disparities can be used to discriminate much finer depth judgments compared to when using absolute disparities (Blakemore, 1970; Westheimer, 1979). Points along the Vieth-Müller circle shown in **Figure 1.1** are the locus of points in space that are located on corresponding points on the retinae and have zero disparity.

Binocular disparities are not limited to the horizontal domain, and vertical disparities also play a role in depth perception. Here, an object with a given height will have a retinal projection that extends further vertically for the side of the object that is closest to the eye,

provided the object is located within a certain proximity (Rogers & Bradshaw, 1993). This relationship will be different for the same object in the other eye, depending on the position and eccentricity of the object in the visual field. Like horizontal disparity, the human visual system is able to compute the difference in these disparities between the two eyes. This cue has been observed to supplement depth estimates from horizontal disparity (Bishop, 1989; Mayhew & Longuet-Higgins, 1982), provided the stimuli is sufficiently large (Bradshaw, Glennerster, & Rogers, 1996) and presented close to the observer (Rogers & Bradshaw, 1993).

1.2.1 Stages involved in binocular perception

There are two important steps in the processing of binocular disparities. First, a feature in the image of one eye needs to be matched with the corresponding feature in the other eye. This is referred to as the correspondence problem (Marr & Poggio, 1979) and is nontrivial as the difference in horizontal position is dependent on the depth of the object. The visual system must therefore exclude false matches that do not correspond to the same feature through a filtering process and identify the one percept that leads to a globally consistent solution (Cumming & Parker, 1997). Second, these small horizontal displacements between the eyes are computed to reveal the depth of the object. However, further processing stages may be required depending on the task of the visual system, such as the computation of relative disparity from two absolute disparities (Cumming & Parker, 1999; Thomas, Cumming, & Parker, 2002). It may also require filling-in of the spatial region between the two locations where disparity was calculated (e.g., to perceive 3D shape) (Backus, Fleet, Parker, & Heeger, 2001; Orban, 2011; Orban, Janssen, & Vogels, 2006b), or for depth segmentation between different surfaces (e.g., to break camouflage) (Bredfeldt & Cumming, 2006).

1.3 The neurophysiology of binocular disparity

Where do these stages of binocular disparity processing take place in the human visual system? The binocular images that fall on the retinae are separated into left and right visual fields in the lateral geniculate nucleus (LGN), but inputs from each eye remains segregated. Although LGN responses for the non-dominant eye of an anaesthetised cat interacted with the activity evoked in the dominant eye (Singer, 1970; Suzuki & Kato, 1966), responses here are not disparity-selective. The observed interaction did not depend on the phase of a sinusoid grating that was shifted between the two eyes (Xue, Ramoa, Carney, & Freeman, 1987), thus neurons in the LGN fired consistently irrelevant of the disparity that was presented. More recently, a neuroimaging study found fMRI responses in the LGN could not reliably discriminate depth position (Preston, Li, Kourtzi, & Welchman, 2008), confirming the lack of binocular responses in the LGN.

Signals from the LGN are then relayed to the primary visual cortex in the occipital lobe, where binocular responses are first combined. Disparity-selective neurons were originally observed in the primary visual cortex of the anaesthetised cat (Barlow, Blakemore, & Pettigrew, 1967; Pettigrew, Nikara, & Bishop, 1968) where neurons responded preferentially to a moving vertical bar for appropriate disparities when presented binocularly within its receptive field. In this section, I will outline what we have learned since this important discovery. Specifically, the different types of responses that have been observed in binocular neurons of the primary visual cortex and how these responses are thought to be used to calculate disparity.

1.3.1 Response profiles of disparity-selective neurons

Sensitivity to binocular disparity was subsequently revealed in the awake, fixating monkey (Poggio & Fischer, 1977; Poggio, Gonzalez, & Krause, 1988), where the authors identified four classes of neurons: Tuned-excitatory neurons that have a narrow, symmetrical profile that responds maximally when presented with a disparity at, or close to, zero. Tuned-inhibitory neurons responded maximally to all disparities except those close to zero. Near-preferred neurons have an excitatory response to all crossed disparities (in front of the horopter) and an inhibitory response to all uncrossed disparities (behind the horopter), while far-preferred neurons are the inverse of this. Poggio (1995) extended this definition to include neurons that are tuned to near or far disparities. However, it was soon revealed that these categories are not discrete; rather, there is a continuum of tuning for disparities that has been observed in both the primary visual cortex (Prince, Cumming, & Parker, 2002) as well as in extrastriate region MT (DeAngelis & Newsome, 1999).

1.3.2 Binocular energy model

The most basic disparity processing is thought to be performed by 'simple cells' of the primary visual cortex. These are orientation-selective and contain subregions within their receptive field that alternate between preferences for bright and dark stimuli, allowing for fine discriminations of position. In contrast, 'complex cells' cannot be divided into subregions and respond to a binocularly-presented stimulus anywhere within its receptive field. Responses of these complex cells have been recreated from the combination of outputs of at least four simple cells, using the binocular energy model (Ohzawa, Deangelis, & Freeman, 1990). Here, responses of pairs of simple cells between the eyes are added together, though mutually inhibited such that a lack of response in one eye cancels the excitation from the other eye.

Pairs of simple cells that are in quadrature phase (i.e., their spatial phases differ by 90 degrees) are combined, and are then half-wave rectified (negative values are discarded) to form the response of the complex cell.

This model can be used to explain the tuning functions observed in disparity-selective neurons. Similarly, through the addition of a simple cell in each eye to threshold certain outputs, a revised model (Read, Parker, & Cumming, 2002) can explain the inversion of the tuning function observed in the primary visual cortex when anticorrelated RDS (where each dot in one eye is spatially matched but contains the opposite contrast to the other eye) are presented (Cumming & Parker, 1997). This model is particularly convenient as the same computation and neural circuitry that can be used to explain the behaviour of complex cells from simple cells can be used to explain the response of relative-disparity-selective neurons from those selective to absolute disparities (Thomas et al., 2002).

1.3.3 Position and phase difference models

There are two ways in which binocular neurons of the primary visual cortex may encode binocular disparity. It could be calculated through an interocular shift in the horizontal position of the receptive field, where the response profile remains similar for both eyes (e.g., Maske, Yamane, & Bishop, 1984). Alternatively, it could be due to a difference in the phase between the two eyes, where receptive field position remains the same and instead the response profile changes (Ohzawa & Freeman, 1986). Both of these hypotheses have been supported empirically. Neurons in the primary visual cortex of the cat showed selectivity towards encoding positional differences in the receptive field in response to a moving vertical bar (Maske et al., 1984; Nikara, Bishop, & Pettigrew, 1968), while phase differences were observed by varying the interocular phase of a sinusoid while also recording from V1 neurons

of the cat (Freeman & Ohzawa, 1990; Ohzawa & Freeman, 1986). More recently, it is thought that both of these processes contribute to disparity calculations (Fleet, Wagner, & Heeger, 1996), although phase differences appear to play a relatively larger role in this process (Anzai, Ohzawa, & Freeman, 1997, 1999).

1.4 Neural responses to disparity in the visual cortex

In the previous section, I outlined how neurons in the primary visual cortex are likely to compute binocular disparities. I will now discuss the role these binocular neurons play in different regions, since responses have been observed throughout the visual cortex (for a review: Gonzalez & Perez, 1998; Neri, 2005; Parker, 2007). In V1, responses are likely to reflect the processing of local disparity signals, as neurons respond equally to stimuli with different depth that appear identical within the receptive field (Cumming & Parker, 2000). Further, disparity-tuning curves vary systematically to anticorrelated RDS which are geometrically matched but do not give rise to stereopsis (Cumming & Parker, 1997).

Interestingly, the functions of disparity-selective neurons change as early as V2. Here, neurons have shown selectivity towards the processing of cyclopean edges (an edge defined only by a change in disparity) (Bredfeldt & Cumming, 2006; Qiu & von der Heydt, 2005; von der Heydt, Zhou, & Friedman, 2000). Similarly, V2 neurons have responded to disparity-induced illusory contours that were beyond the size of their individual receptive field (Bakin, Nakayama, & Gilbert, 2000), suggesting a role in the assignment of stereoscopic boundaries and the segregation of surfaces. The functions of binocular neurons in later visual areas differ depending on the visual pathway, and this is discussed more thoroughly in the next section.

1.4.1 Distinctions in the processing between dorsal and ventral visual pathways

The visual cortex is divided into two, anatomically separated pathways: the ventral stream which passes information from the primary visual cortex to the inferior temporal lobe and is involved in visual perception and object identification, and the dorsal stream projects information from the primary visual cortex to the posterior parietal cortex and is involved in the guidance of actions (Goodale & Milner, 1992). It is reasonable to think that since the visual streams play very different roles in the processing of vision, the processing of binocular disparities would also differ between the visual pathways.

Tyler (1990) proposed that such a division originates in the LGN and proceeds to these later regions, with neurons in the magnocellular layer performing a rapid computation of coarse (large) disparities and neurons in the parvocellular layer used for calculation of fine (small) disparities, since responses here had a slow response latency and were improved for decisions requiring high precision. However, subsequent findings revealed that many neurons in these layers converge in the primary visual cortex and that these layers are not clearly separated between the two pathways (e.g., Merigan & Maunsell, 1993; Nassi, Lyon, & Callaway, 2006).

Although this particular theory was largely abandoned, the suggestion of separating the visual pathways by coarse and fine disparities has received further empirical support. To examine the processing of coarse disparities, neuronal responses were recorded while monkeys discriminated depth position of a central plane that contained a large disparity, while the proportion of dots with disparity noise (dots that were correlated but located at random disparities) was varied. Similarly, the processing of fine disparities was examined through a similar task where monkeys performed a relative disparity judgment between much smaller disparities in the absence of noise. Responses of individual neurons in MT were typically as

sensitive as that of the monkey for the coarse task (Uka & DeAngelis, 2003, 2004). Similarly, electrical microstimulation (which forcibly activates nearby neurons) of MT neurons was found to bias perceptual judgments on this task (DeAngelis, Cumming, & Newsome, 1998; Uka & DeAngelis, 2006). In contrast, neurons at the end stages of the ventral pathway responded to fine changes in depth position (Janssen, Vogels, & Orban, 2000b; Uka, Tanabe, Watanabe, & Fujita, 2005) and performance on the fine task was affected by microstimulation in ventral region V4 (Shiozaki, Tanabe, Doi, & Fujita, 2012). Importantly, microstimulation of dorsal MT neurons was *not* found to affect behavioural performance for the fine task (Uka & DeAngelis, 2006), thereby providing evidence of a dissociation between the visual pathways. A further neuroimaging study examined fMRI responses to a disparity-defined sinusoid that was presented at one of several pedestal disparities (Minini, Parker, & Bridge, 2010). The authors found that dorsal regions responded differentially to large changes in the pedestal disparity and suggested the dorsal visual pathway may be more coarsely-tuned, providing greater depth discrimination for coarse disparities.

Alternatively, it has been suggested that the processing of absolute and relative disparities is a key parameter in separating the visual streams (Neri, Bridge, & Heeger, 2004; Uka & DeAngelis, 2006; Umeda, Tanabe, & Fujita, 2007), with results from the coarse and fine tasks being attributed to the processing of absolute and relative disparities, respectively. However, as described in the following section, responses to absolute and relative disparities have been observed across both pathways, suggesting this distinction is less appropriate.

1.4.2 Absolute and relative disparities

Binocular depth perception is more likely to rely on relative, rather than absolute, disparity signals. One likely reason for this is the fact that relative disparity is calculated independent of

eye position, so using relative disparities would allow stereopsis to be observed throughout the visual scene. Absolute disparities are unable to explain our ability to make fine depth judgments, as psychometric thresholds are vastly improved for stimuli that are presented alongside a nearby reference (Westheimer, 1979; Westheimer & McKee, 1978). Further, large changes in absolute disparity are not always reflected with a change in depth perception (Erkelens & Collewijn, 1985; Regan, Erkelens, & Collewijn, 1986). It should be noted, however, that the visual system uses binocular disparities for several purposes other than stereopsis. For example, absolute disparities have been observed to play a role in vergence eye movements and in spatial navigation. In comparison, relative disparities are necessary for making fine depth judgments about the 3D shape of an object.

Responses to absolute and relative disparities are observed in both visual pathways; however this is not the case in the primary visual cortex. Cumming and Parker (1999) used a feedback loop (the absolute disparity of the entire binocular field was altered while controlling for vergence angle and keeping the relative disparity constant) to reveal a clear preference of V1 neurons in the processing of absolute disparities. For one popular experimental paradigm, disparity-tuning curves were generated by measuring the neuronal response of several absolute disparities, while keeping the depth of the surround constant. This was repeated for multiple surround positions, such that a shift in the disparity-tuning curve reflected processing that was invariant of absolute depth position, revealing selectivity towards relative disparity processing. This method was used in several extrastriate regions, and revealed that a small proportion of neurons in V2 responded to relative disparity (Thomas et al., 2002). A larger proportion was evident in ventral region V4 (Umeda et al., 2007) while relative disparity sensitive neurons were generally not observed in dorsal regions V3, V3A and MT (Anzai, Chowdhury, & Deangelis, 2011; Uka & DeAngelis, 2006). Interestingly, this

result changed significantly to a change in stimulus configuration. Instead of a centre-surround configuration, Krug and Parker (2011) varied the depth position and relationship between two superimposed, transparent planes and found a much greater shift of the tuning curve in the responses of MT neurons, highlighting the importance of stimulus configuration when examining relative disparity signals.

Neuroimaging studies have provided conflicting results on the issue. One study examined selectivity for absolute and relative disparities in humans by using an fMRI paradigm that showed blocks of stimuli where either the absolute or relative disparity was held constant. The level of adaptation in the BOLD signal was compared to paired blocks where the absolute or relative disparity varied between trials. The authors found that dorsal regions responded primarily to absolute disparities, while ventral regions responded with similar sensitivity to both absolute and relative disparities (Neri et al., 2004). In comparison, Cottereau *et al.* (2011, 2012a) took a novel approach by measuring the temporal dynamics of disparity processing using EEG. In these studies they compared responses to a centre-surround stimulus when the surround was composed of uncorrelated noise (absolute disparity) versus a correlated surround (relative disparity). They found significant responses across extrastriate areas, with increased sensitivity to relative disparity in LO and V3A, and a change in phase in regions V4, V3A and hMT+/V5 which they attribute to the neural calculation of the relative disparity. Further, it was only dorsal region V3A that modulated its response to changes in the position of the surround. Thus, evidence is converging from both neurophysiology and neuroimaging methods to suggest that relative disparity processing in fact takes place across both visual pathways. In line with traditional characterisation of cortical visual processing pathways, it is likely that this serves different functions between the

visual streams: in the navigation and segmentation of moving objects in the dorsal stream and for 3D shape perception in the ventral stream.

1.4.3 Neural responses to different surface configurations

The vast majority of the studies mentioned thus far have used vertical bars, sinusoidal gratings or random-dot stereograms of frontoparallel surfaces. However, objects in the real world are composed of complicated 3D shapes that include slants and curvature. Moreover, it is likely that neural responses depend on the type of surface configuration used, since the ventral visual pathway appears to be more specialised for the processing of 3D shapes and for object recognition (Kourtzi & Kanwisher, 2000, 2001).

Neuroimaging responses to flat planar surfaces placed in depth have repeatedly shown high sensitivity for dorsal regions, particularly V3A. Here, fMRI responses were significantly higher for planes that contained nonzero disparities in comparison to when it was presented at the fixation plane (Backus et al., 2001). Similarly, fMRI responses in both monkeys and humans revealed intermediate and later dorsal regions were most sensitive to a disparity-defined checkerboard that contained squares located at several depths compared to a single plane at zero disparity (Tsao et al., 2003). This preference of V3A and other dorsal regions to planar stimuli was also observed in studies comparing responses between correlated RDS and an anticorrelated RDS (Preston et al., 2008) or an uncorrelated RDS (Cottereau et al., 2011, 2012a).

In comparison, responses to disparity gradients (i.e., disparity-defined slants) are seen in relatively later regions of the visual pathways, consistent with a transition towards higher-order selectivity. For example, one fMRI study measured the level of neural adaptation to repeated presentations of disparity and perspective cues to slant (Welchman, Deubelius,

Conrad, Bulthoff, & Kourtzi, 2005). The authors found LO and hMT+/V5 responded to both disparity and perspective cues and appeared to represent the perceived slant. This supports several studies that measured neuronal responses in different visual areas during presentation of stereograms that were rotated in-depth along either the horizontal or vertical axis (to produce slanted or tilted surfaces, respectively). Roughly half of neurons in MT were tuned to slants or tilts, and responded independently of one another (Nguyenkim & DeAngelis, 2003). However, these responses were modest in comparison to those observed in later dorsal areas: the vast majority of measured neurons in parietal regions CIP and AIP responded to disparity gradients (Shikata, Tanaka, Nakamura, Taira, & Sakata, 1996; Srivastava, Orban, De Maziere, & Janssen, 2009; Taira, Tsutsui, Jiang, Yara, & Sakata, 2000), and reversible inactivation of CIP neurons was found to impair the monkey's ability to discriminate disparity-defined slant on half of their attempts (Tsutsui, Jiang, Yara, Sakata, & Taira, 2001). For regions in the ventral stream, V4 neurons showed only mild selectivity towards disparity gradients (Hegde & Van Essen, 2005) while neurons in TEs of the inferotemporal cortex showed some selectivity towards slant, but many more responded significantly more to 3D curvature (Janssen et al., 2000b).

A similar pattern was observed for surfaces that were curved in depth. Single-unit studies showed a large proportion of neurons sensitive to 3D curvature in TEs (Janssen, Vogels, Liu, & Orban, 2003; Janssen, Vogels, & Orban, 2000a; 2000b), LIP (Durand et al., 2007) and AIP (Durand, Peeters, Norman, Todd, & Orban, 2009; Srivastava et al., 2009; Theys, Srivastava, van Loon, Goffin, & Janssen, 2012b). In contrast, fMRI results showed a specialisation for 3D curvature throughout dorsal regions, beginning in V3A and observed through to the endpoint of the dorsal stream, phAIP (Georgieva, Peeters, Kolster, Todd, & Orban, 2009). Interestingly, reducing the coherence in disparity-defined 3D shapes was found

to similarly reduce the fMRI signal in LO, V3 and hMT+/V5 (Chandrasekaran, Canon, Dahmen, Kourtzi, & Welchman, 2007), though it should be noted that this is the only study to parametrically introduce disparity noise into the stimulus.

Importantly, the selectivity towards 3D shape is also likely to have different purposes between the visual streams. Neural responses to a curved surface were observed before the perceptual decision was made in ventral region TEs, and after the perceptual decision in dorsal region AIP (Verhoef, Vogels, & Janssen, 2010). The authors suggested that 3D shape processing is likely performed in the temporal cortex for perception of the surface, while responses in the parietal cortex are used to emphasise the role of grasping movements.

1.5 Linking neural activity to behaviour

As illustrated in the previous section, neurons sensitive to binocular disparity have been observed throughout the visual cortex. However, this does not directly imply a role in disparity processing, as disparity-selective neurons have also been observed to play a role in a variety of other visual tasks including figure-ground segregation (Qiu & von der Heydt, 2005) and vergence eye-movements (Takemura, Inoue, Kawano, Quaia, & Miles, 2001). Further, it does not reveal whether these areas have information regarding the binocular depth percept. Simply measuring neural responses to disparity stimuli will not answer this question. Instead, it is necessary to directly manipulate the binocular depth percept while simultaneously measuring the neural response.

Little evidence has suggested that V1 neurons show processing that reflects depth perception. In a depth discrimination task where monkeys had to judge between different magnitudes of near and far disparities, some neurons showed sensitivity that matched the behavioural performance of the monkey, and on average the sensitivity of the measured

neurons was slightly worse (Prince, Pointon, Cumming, & Parker, 2000). However, the authors discussed that stimulus intervals were not appropriately spaced and re-measuring psychometric thresholds (without simultaneous neuronal recordings) improved the thresholds by a factor of four. Further, V1 neurons have been observed to fire consistently to a sinusoid that was horizontally shifted one period of phase for the measured receptive field in one eye, such that local disparity signals were identical but the overall depth percept was changed (Cumming & Parker, 2000). V1 neurons have also displayed tuning functions to anticorrelated RDS (though inverted to those observed with correlated RDS), even though these do not produce a sensation of depth (Cumming & Parker, 1997). This finding is supported by neuroimaging evidence, where similar decoding accuracies were observed between correlated and anticorrelated RDS in early visual areas (Preston et al., 2008). These studies suggest that single V1 signals are unable to account for depth perception and that it requires further processing in extrastriate regions.

One approach to examine regions whose activity corresponds with perception is to compare the responses of correlated and anticorrelated RDS. Here, the overall properties of the stimulus remain the same, but depth is not perceived for anticorrelated RDS as the visual system is unable to find a solution that is globally consistent (Julesz, 1971). Single-unit studies revealed a reduction in the response of neurons to anticorrelated RDS by V4 (Kumano, Tanabe, & Fujita, 2008; Tanabe, Umeda, & Fujita, 2004) and by the stage of the inferotemporal cortex, disparity-sensitive neurons had completely attenuated their response to anticorrelated RDS (Janssen et al., 2003). This was also observed at the endpoint of the dorsal stream, where neurons in the anterior intraparietal areas (AIP) also suppressed activity during presentation of anticorrelated RDS (Theys et al., 2012b), revealing that both visual pathways contain information that potentially reflects the binocular depth percept.

It is uncertain where information regarding the depth percept is available in the dorsal visual pathway. Preston *et al.* (2008) used fMRI and multivariate techniques to find a preference towards correlated RDS in all measured dorsal regions (as well as in ventral region LO), suggesting this occurs at an earlier stage of the dorsal pathway. However, this conflicts with evidence found using structure-from-motion stimuli. Here, 3D rotating cylinders are portrayed using two planes of dots that move in opposite directions. When all dots have the same binocular disparity, the direction of rotation is inherently ambiguous and the perceived direction of rotation alternates over time. However, when these dots are given binocular disparities, the direction of rotation is defined unambiguously. Interestingly, many MT neurons showed a significant change in activity to an ambiguous structure-from-motion cylinder depending on the direction of rotation that was perceived, and this was consistent with the preference of the neuron when the cylinder contained binocular disparities (Bradley, Chang, & Andersen, 1998). Further, using choice probabilities (the probability that the perceptual decision of the monkey can be predicted from the firing rate of the neuron; Britten, Newsome, Shadlen, Celebrini, & Movshon, 1996), trial-to-trial variability in the responses of MT neurons to the structure-from-motion stimulus was revealed to be strongly correlated to the reported percept of the monkey (Dodd, Krug, Cumming, & Parker, 2001), suggesting an involvement of these neurons in the making of the perceptual decision. However, these neurons do not yet underlie stereoscopic depth, as neurons with these high choice probabilities in the rotating cylinder task also responded when anticorrelated dots were used and no depth percept was available (Krug, Cumming, & Parker, 2004).

From this body of evidence, and combined with studies that examined sensitivity of individual neurons in relation to the monkey's performance on a coarse disparity task (see *1.4.1 Distinctions in the processing between dorsal and ventral visual pathways*), it is unclear

whether responses in MT contain information about the binocular depth percept. In comparison, responses in the parietal cortex do appear to reflect the stimulus percept. The activity of LIP neurons could be used to predict the perceptual decision of a monkey. This was observed during both a left-right motion discrimination task (Shadlen & Newsome, 1996) and also on an apparent motion task where the direction of motion is entirely ambiguous and perceptually bi-stable (Williams, Elfar, Eskandar, Toth, & Assad, 2003).

The discussed experiments in this section had one main similarity: they all examined neural responses while the perception of the stimulus was manipulated directly. They revealed a clear pattern that signals responding to binocular depth perception are observed in both later ventral and dorsal visual pathways, though the specific cortical sites where this transformation takes place is not yet known. Further, it is consistently shown throughout this review that the role the two visual pathways appear to use information about binocular depth differently. It seems the main purpose of binocular depth in the dorsal stream is to aid navigation: it is used to extend and segregate surfaces, and guides movements of both the hand (e.g., grasping, reaching) and the eyes (e.g., vergence). In comparison, the ventral stream seems to be involved with perception, where the three-dimensional shape of the object is formed and object recognition takes place.

1.6 Overview of chapters

As evident from this review, in the past 50 years since the discovery of disparity-selective neurons in the visual cortex, we have learned a considerable amount regarding the cortical processing of binocular disparity. However, this is a complicated process and is even more challenging to understand through the several different purposes of disparity in the visual system. In this thesis, I address several core issues relating to the processing of binocular

disparity in the human visual system: I investigate the relationship between cortical activity and perception, the significance of the reference plane on depth configurations, and the topography of disparity signals on the cortical surface.

Chapter 2. In the general methods, I provide background on the main apparatus and procedures used in the experimental chapters. I also discuss some of the major methodological issues that need to be considered when using these approaches.

Chapter 3. This experimental chapter examines the relationship between cortical activity and the perception of depth. In particular, I examined whether disparity signals between the ventral and dorsal visual pathway could be separated through the processing of signal-in-noise and feature difference tasks. I show that regions in the later dorsal visual cortex were closely related to perception in both cases.

Chapter 4. Closely related to Chapter 3, I used anticorrelated random-dot stereograms to investigate cortical regions that had solved the correspondence problem and whose responses were consistent with the depth percept of the observer. The results again highlight the role of later dorsal regions that had solved the correspondence problem and were observed to mirror binocular depth perception.

Chapter 5. For this chapter, I examine the role of stimulus configuration on the processing of disparity. As evident in several places through my review of the literature (this chapter), a different stimulus configuration can have significant implications for the findings of an experiment. In Chapter 5 I therefore investigate the importance of the stimulus reference plane. I find early visual areas calculate only local changes in disparity while intermediate dorsal regions can potentially compute disparity with regard to more than one reference.

Chapter 6. Here, we attempted to identify structure in the representation of binocular disparity in the visual cortex. We estimated the disparity preference of each voxel in two distinct ways, and displayed these preferences on a flatmap of the cortical surface. Although we did not observe a topographic map of disparity, we observed a large cluster of voxels within V3A, V3B/KO and V7 that consistently showed bias towards crossed disparities, and those of a greater magnitude.

Chapter 7. In the final chapter, I review the findings of the previous chapters and highlight what these results contribute to our understanding to the cortical responses of binocular disparity.

2. General Methods

A broadly similar method underlies all experiments contained in this thesis. They all use random-dot stereograms to present depth planes to healthy adult observers inside an fMRI scanner at the Birmingham University Imaging Centre. A multi-voxel pattern classifier was used in the analysis of fRMI data throughout all experimental chapters. The precise details of these techniques differ by experiment and will be explained in each chapter. Instead, here I provide some background on these methods and highlight the benefits of these approaches. I will also discuss their limitations and where appropriate, the measures we took to control for them.

2.1 Participant recruitment, screening and ethics

All participants who took part were staff and students of the University of Birmingham. Whenever necessary, we recruited participants through advertisements located around the university or on the university's job recruitment website. All participants gave written informed consent prior to participation and were paid for their time. We first screened participants to identify a possible stereo deficit or any contraindications to fMRI scanning. If suitable, we then screened participants for their stereoacuity in the laboratory haploscope to ensure they could perceive disparities as small as 1 arcmin. This was conducted using a custom-made stereo-test and took place immediately prior to experiment training. All projects were evaluated and given ethical approval by the University of Birmingham Ethics Committee.

2.2 Choice and design of stimuli

We used random-dot stereograms to present depth planes to observers for all of our experiments (Julesz, 1971). Each stereogram consists of an array of random dots correlated between the two eyes, where a small horizontal shift to a central patch of dots is made between the images of the two eyes. Depth is only perceived after binocular fusion has taken place, thereby isolating processing to binocular cues. Monocular cues (e.g., texture, linear perspective, motion parallax) are still present in the stimulus, but all signal a flat display and are therefore uninformative to indicate the correct depth. The direction and magnitude of this horizontal shift affects the resulting perceived disparity (**Figure 2.1a**).

2.2.1 Generating random dot stereograms

Stimuli were created in C# using custom-made functions that called OpenGL graphics libraries and were rendered using anti-aliasing. In all experiments, the stereogram consisted of two parts: a critical test plane and a surrounding region (the shape of these differed between experiments). This surround was necessary to disguise the horizontal shift of the test plane disparity and was used to keep the size of the stereogram constant for all trials. Dots were generated in black and white with an equal proportion of both, such that the mean luminance of the stimulus matched that of the background (mid-grey). Dots were given a Gaussian luminance profile (0.15° diameter at half-height) to soften edges. Similarly, the background consisted of coarsely distributed black and white squares (0.9° side length; 75% density) so that the stimulus was surrounded by a low spatial frequency pattern at zero disparity.

2.2.2 Controlling for monocular zones

The horizontal shift used to create a disparity in the stimulus leaves a corresponding empty space on the other side of the stimulus (striped region, **Figure 2.1c**). This is known as a monocular zone since this space only contains dots in the image of the opposite eye. This empty space is a monocular cue where the location and width of the space varies systematically with the disparity presented and has the potential to influence fMRI measurements without stereopsis taking place. Furthermore, early visual areas are particularly sensitive to contours and edges (Felleman & Van Essen, 1991; Grill-Spector & Malach, 2004) and therefore fMRI responses may relate to the contours of the monocular zone rather than the presented disparity.

Depending on the experiment, we addressed this concern in one of three ways. First, the surround was filled with correlated dots and once the test plane was shifted, it created a monocular zone and a region of overlap with the surround that were of equal size. We therefore filled the monocular zones with the overlapped dots from the opposite side of the test plane (**Figure 2.1c**). These dots, once their position was changed, no longer have a corresponding dot in the other eye and were monocular (uncorrelated); a normal consequence of stereoscopic viewing (Nakayama & Shimojo, 1990). This is illustrated by **Figure 2.1b** where, between the eyes, a surface that is nearer to the observer occludes different parts of the surface behind it, such that these regions are only visible to one of the eyes. We filled in monocular zones using this method for the signal-in-noise task in Chapter 3 and in Chapter 4. One consideration with this method is that with increasing disparity magnitudes, the width of the test plane is correspondingly reduced. However, the disparities we used for these experiments (±6 arcmin) were sufficiently small that any such size difference was not noticeable. In Chapter 6, we used a similar method where we created two surrounds that were

filled with monocular dots only once the disparity shift was calculated. This kept the size of the test plane constant (**Figure 2.1d**) and was particularly beneficial for disparity planes that were irregularly-shaped and were unable to be moved from one side of the test plane to the other. These solutions were elegant as all monocular cues were removed and the density of dots was preserved throughout the entire stereogram. However, if both central and surround planes contained nonzero disparities these monocular dots generally appeared at a third depth; at the plane of the screen. In these cases, our solution was to use a low density of dots so that the monocular zones and their edges were hidden among the sparser dot distribution. We used this approach in Chapter 3 for the feature difference task and in Chapter 5, since the stimuli in these experiments contained nonzero disparities for both centre and surround planes.

2.2.3 Preventing adaptation

Repeated presentation of a particular disparity has been found to create adaptation and cause the perception of subsequent presentation of that (or similar) disparities to be biased in the opposite direction (Blakemore & Hague, 1972; Blakemore & Julesz, 1971; Stevenson, Cormack, Schor, & Tyler, 1992). This was observed even on random-dot stereograms where monocular cues provide no information on the adapting stimulus, suggesting that this is a global process that takes place after stereopsis. Neurally, repeated stimulus presentation causes a reduction in fMRI activity in contrast to when a novel stimulus is displayed (Buckner et al., 1998) and this difference has been successfully exploited to differentiate between types of disparity processing (absolute versus relative disparities) in the human visual cortex (Neri et al., 2004). We therefore need to consider the effect of adaptation in our fMRI studies, especially since repeated stimulus presentations are needed to generate clearer responses due to the slow hemodynamic response of the BOLD signal (see *2.5.1 The BOLD signal*). To

address this concern in our experiments, we first separated stimulus presentations with an inter-stimulus interval to reduce build-up of adaptation. We then introduced small variations in the disparity of the random-dot stereogram between trials for the majority of our experiments. Generally, we allowed the disparity to randomly vary up to ±1 arcmin on each trial, but in Chapter 6 we rotated the stimulus slightly so disparity edges were not located continuously in the same position.

2.2.4 Preventing eye-movements

The presentation of disparities can evoke eye-movements away from the plane of fixation at ultrashort latencies (Busettini, Miles, & Krauzlis, 1996; Masson, Busettini, & Miles, 1997); less than the length of stimulus presentation that we used in our experiments. As such, we took steps to minimise this confound and ensure our results were due to the presentation of disparities and not to potential eye-movements. In all tasks a fixation marker was presented at the centre of the stimulus to control the eye-position across trials. This fixation marker contained both horizontal and vertical nonius lines that were split between the images of the two eyes and when combined looked like a cross-hair. Further, stimulus dots were not presented immediately surrounding the fixation marker to ensure participants could easily recognise if the nonius lines were misaligned and therefore vergence had shifted away from the plane of fixation. Participants were instructed to ensure the nonius lines were in alignment at all times throughout the experiment. We also surrounded the stimulus with a zero-disparity reference of black and white squares that were used to further encourage vergence at the fixation plane.

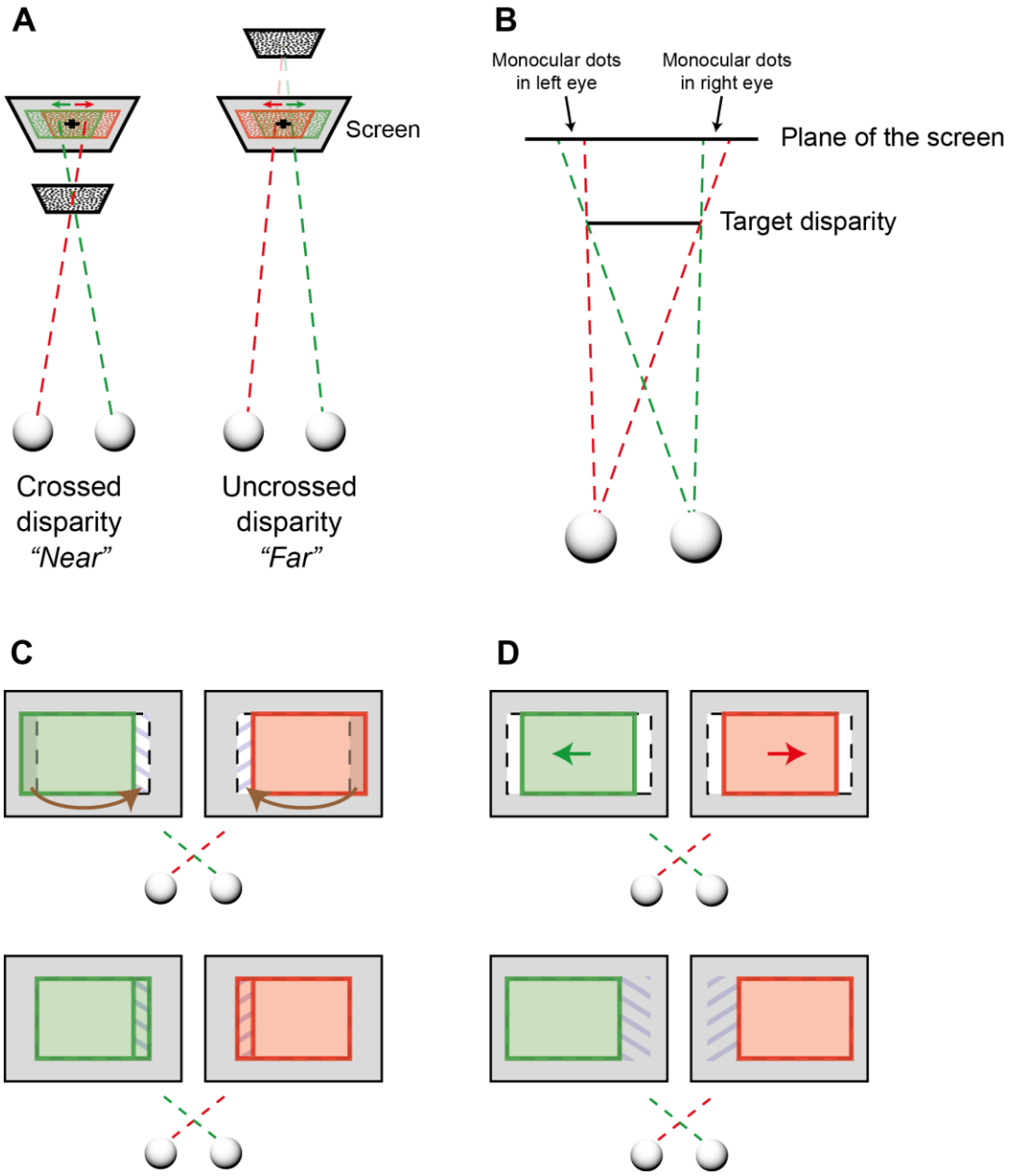


Figure 2.1: Depth from random-dot stereograms and the issue of monocular zones. (A) An aerial view of two random-dot stereograms where the dots in each eye are horizontally displaced on a screen to give the impression of stereoscopic depth. Each colour represents the unique view of that eye and each line is the optic path to the centre of the stereogram visible to that eye (fixation is always at the marker at the centre of the screen and does not change depth). The direction of these shifts determines whether crossed or uncrossed disparities are perceived. (B) An aerial view of da Vinci stereopsis, where between the two eyes the target plane occludes different parts of a surface behind it, thereby creating monocular zones. (C) One method to remove monocular zones in our stimuli. The shift of the central plane creates a monocular zone (blue stripes, top image) in each eye. The grey region is the surround plane of the RDS. The dotted line is the central patch of dots before being shifted to create binocular disparities. The green and red rectangles represent the central plane of the RDS to each corresponding eye. The region of dots that overlap with the surround in the top image are moved to fill the monocular zone. The bottom image shows this once this region has been moved. (D) A second method to remove monocular zones. Here the surround does not border the central plane until the horizontal shift takes place, where the monocular zones are then filled with dots at the depth of the surround.

For experiments in Chapters 5 and 6, participants completed a vernier task (Popple, Smallman, & Findlay, 1998) throughout all scans to direct attention to the fixation plane and ensure correct vergence alignment. Here, a vertical line was presented inside fixation at the same time as stimulus presentation and contained an offset in its horizontal position relative to the vertical nonius line located immediately above. Participants were asked whether the vernier stimulus was to the left or right of the nonius line. Both lines were presented monocularly to separate eyes such that correct vergence was required to correctly identify the relative position of the vernier stimulus, whereas a shift in the results indicated a change in vergence angle away from fixation. This task was not used in Chapters 3 and 4 as we were interested in measuring the depth percept of the observer simultaneous to recording fMRI measurements. However, for one experiment (the feature difference task) in Chapter 3 that used nonzero disparities for both centre and surround target planes, participants completed a self-paced version of the vernier task prior to the start of fMRI scanning. This was to ensure that no particular condition elicited an eye-movement away from fixation.

Finally, for experiments in Chapters 3 – 5, we recorded horizontal eye-movements for a subset of the participants using a monocular limbus eye-tracker (CRS ltd) in separate fMRI sessions. We used a custom-built mount to attach INFITEC spectacles (see *2.4.1 Spectral filters and dual projector system)* to the head coil, as well as to hold an infrared sensor beneath the spectacles and in front of the participant's right eye. This sensor was connected via fibre-optic cable to an eye-tracker located in the scanner control room. Horizontal eye position was calibrated before each run and then participants completed the experimental task as previously trained. Custom-made software created in MATLAB was used to discriminate between saccades, blinks and periods of fixation.

2.3 Laboratory equipment for stimulus display

Prior to scanning, all participants underwent psychophysical training on our laboratory haploscope at the University of Birmingham. This was to ensure disparities for the specific experiment could be accurately perceived and to give participants prior exposure to the stimuli before participating in the fMRI sessions. Our haploscope was a modified Wheatstone stereoscope (Wheatstone, 1838), where two mirrors at right angles were positioned close to the eyes, such that each eye was presented with an image from a different CRT display. This allowed us to independently manipulate the images presented to each eye. We used 21-inch CRT displays (ViewSonic FB225f) with a refresh rate of 100 Hz. Vergence angle was controlled by hinging both the mirror and CRT display about a point underneath each eye. Inter-pupillary distance was configured by adjusting the separation between these hinges. We generated binocular disparities by shifting the horizontal position of dots in the stereograms presented on each CRT display.

2.4 fMRI display equipment

Stimuli were projected images from dual projectors and presented to participants on a translucent screen located inside the fMRI scanner bore. Participants viewed the screen (optical path length, 65 cm) through a mirror positioned on the head coil, angled at 45°. The graphics card (Quadro FX 4400) generated stimuli using 1280 × 1024 pixels at a refresh rate of 60 Hz. Stimuli were controlled using MATLAB 2009a with the PsychToolBox 3 component installed (Brainard, 1997; Pelli, 1997). Participants responded to visual stimuli through an MR-safe button box held in their dominant hand.

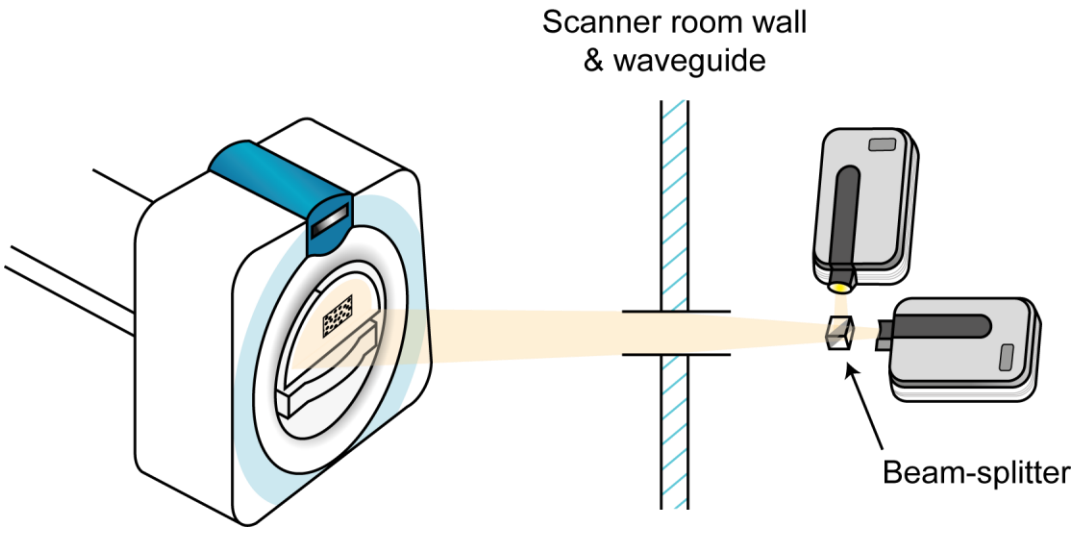


Figure 2.2: Dual projector setup. The projected images are transmitted from two projectors and become overlaid via a beam-splitter before being passed through a waveguide and onto a translucent screen at the back of the scanner bore.

2.4.1 Spectral filters and dual projector system

During scanning, participants viewed stereoscopic stimuli through a pair of video projectors (JVC D-ILA SX21) whose images became overlaid via a beam-splitter cube before being passed through a waveguide into the scanner room (**Figure 2.2**). Each projector was fitted with an INFITEC interference filter, which is a band pass filter that divides the wavelengths for each primary colour between the two projectors. Since the INFITEC filters produced negligible overlap between the spectra for each projector, there was minimal crosstalk between the projected images for an observer wearing a pair of corresponding filters. Although these filters allowed clear presentation of stereoscopic images to participants in the fMRI scanner, the necessary use of spectacles limited our ability to measure eye movements.

The beam-splitter was a cube consisting of two glass prisms that were used to divide incoming light into two separate parts. We positioned the beam-splitter at the intersection of the optical paths of the two projectors, such that one part of each projection was transmitted through the waveguide and into the scanner room. The use of a waveguide located in the wall

of the scanner room allowed presentation of the projected images directly onto a translucent screen in the bore of the magnet, without distortion from having to pass through a Faraday cage. It was critical to have the images perfectly superimposed before being optically combined by the beam-splitter. Slight horizontal misalignments between the projectors would result in the presentation of unintended disparities. Therefore, before data acquisition in every fMRI scan, we presented the identical image of a grid using both projectors and manually adjusted the physical position of one of the projectors. Its position could be shifted horizontally, or tilted with respect to the left or right sides of the image. We continued making adjustments until this grid was superimposed and we were perceptually unable to identify a difference in the position of the grid over the region where stimuli were presented.

2.4.2 Luminance calibration

To ensure the image intensity of the projectors remained comparatively similar over time, we routinely examined the luminance output of each projector through one of two measures. First, we used a subjective flicker photometry task where a large circle was displayed, flickering, with its input alternating between the two projectors at 10 Hz. The greyscale value in one of the projectors was fixed at one of 32 linearly-sampled greyscale values, while the greyscale value in the other projector was generated randomly. The observer was asked to increase or decrease this greyscale value until the perceptual brightness appeared matched. This data was linearly fit and the gamma function of the brighter projector was reduced by the slope of this function (e.g., **Figure 2.3**) and then normalised. Second, we used a photometer with fibre-optic cable to measure the candela output of each projector at the location of the screen for different greyscale values. We generated a normalised gamma table for the brighter projector where the highest greyscale value was restricted to the maximum value obtained in

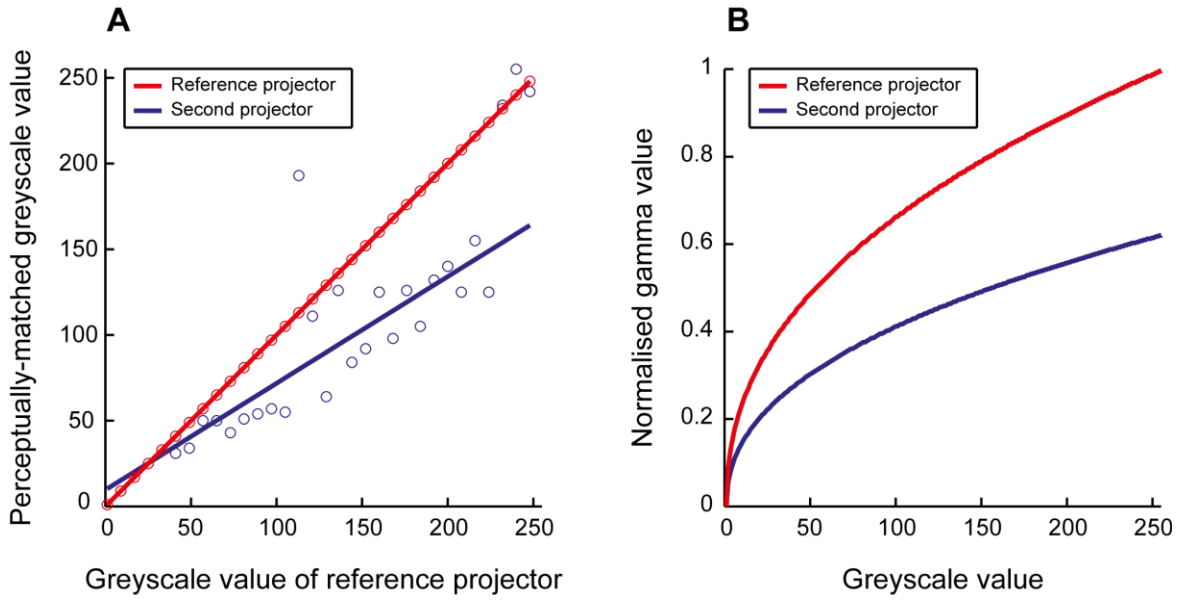


Figure 2.3: Sample results of the flicker photometry task. (A) Greyscale values were flickered between the two projectors. Observers judged the point of isoluminance by adjusting the greyscale value of one of the projectors to match that of the other (reference projector). No task was performed on the reference projector and the perceptual results were assumed to perfectly match the output greyscale values. Least-squares fitting were applied to the data points for each projector. (B) The gamma function of the reference projector (red line) and second projector (blue line), where the latter is scaled by the slope observed by the linear fit in (A).

the weaker projector. The normalised gamma functions for both projectors were implemented into the display using PsychToolBox immediately prior to stimulus presentation.

2.5 fMRI data acquisition

All fMRI scans were performed in a 3 Tesla Philips Achieva MRI scanner at the Birmingham University Imaging Centre. Echo-planar imaging (EPI) data was acquired from blood oxygenation level-dependant (BOLD) signals using an 8-channel SENSE head coil. One experimental session in the scanner lasted approximately 90 minutes and usually contained 8 – 10 runs. ROIs were identified using functional localisers obtained independently from a separate fMRI session. We identified visual areas in the early visual cortex (V1, V2), ventral visual pathway (V3v, V4, LO) and dorsal visual pathway (V3, V3A, V3B/KO, V7,

hMT+/V5). We confirmed the location identified with these localiser scans with previously observed Talairach coordinates to ensure accurate identification of regions.

2.5.1 The BOLD signal

Although fMRI techniques are now commonplace in research around the world, it is important that we review the type of signal that is measured and the issues we need to consider when interpreting fMRI results. We note that fMRI does not measure neuronal activity directly and is instead measuring changes in the metabolic demands of neurons. Specifically, it uses a magnetic field to examine changes in the oxygenation of the bloodstream, a process known as blood oxygenation level dependent (BOLD) imaging. When neural activity increases, there is an increased demand for oxygen which is delivered by oxyhemoglobin and is diamagnetic. Once absorbed, the blood contains deoxyhemoglobin and is paramagnetic. Application of a radio-frequency (RF) excitation by a transmit-receive coil inside the magnetic field aligns the nuclei of hydrogen atoms in deoxyhemoglobin. The direction of rotation of the nuclei is altered by the gradient coils, and allows slices to be oriented in any arbitrary direction. These hydrogen atoms emit energy at the same RF with each pulse and then gradually return to equilibrium. This energy can be measured with the transmit-receiver coil. Once identified, these frequencies are converted from k-space into fMRI images using Fourier transforms.

For our experiments, we used echo-planar imaging (EPI). This uses rapid gradient switching to measure all the phase-encoding steps within a single pulse and reduces the time taken to obtain fMRI images. We used an echo-time (TE) of ~34 ms which was the time from the centre of the RF excitation pulse to the peak of the echo where the signal is measured. We used a repetition time (TR) of either 1500 or 2000 ms depending on the experiment and this

corresponds to time between successive RF excitation pulses. For anatomical scans, we collected T1-weighted images where the time taken for relaxation of the nuclei was longer than the TR. Since different tissues have different relaxation rates, only tissue without fluid (i.e., cortical matter) had enough time to recover from the previous excitation and was presented clearly in the fMRI image. For functional scans, we collected T2-weighted images where the time taken for the nuclei to go out of phase with each other was shorter than the TE, such that only signals from tissue with fluid (i.e., CSF, blood) were presented in the fMRI image.

Neural activity observed indirectly through the BOLD signal in these functional images is modelled by the hemodynamic response function (HRF), and consists of three stages. First, there is a small decrease in the BOLD signal below baseline, reflecting the initial stage of oxygen consumption. Second, there is a large increase in BOLD signal where the increased metabolic demand results in an overcompensation of supply in oxygenated blood. Finally, there is a decrease back to below baseline until the blood volume returns to normal.

There are a number of considerations when interpreting fMRI results. First, the origin of the BOLD signal is not completely understood. It is widely considered to be related to changes in neural activity (Arthurs & Boniface, 2002; Heeger & Ress, 2002) and simultaneous acquisition of neurophysiological and fMRI data has further linked these as related processes (Logothetis & Pfeuffer, 2004). However, it is not certain whether these signals arise from action potentials or synaptic activity and as an indirect measure, is potentially influenced by changes in the body that are unrelated to neural activity. Further, the use of the HRF model makes several assumptions on the relationship between neural activity and the metabolic response, and is not suitable for all circumstances (Heeger & Ress, 2002). Second, statistical controls need to be in place to account for the large number of voxels

measured using fMRI and the high risk of Type I errors (false positives). For our experiments, we used a region-of-interest (ROI) approach that pre-defines selected visual areas through an independent dataset in the localiser scans. Third, due to the delay of the metabolic system in responding to neural activity and the gradual decrease back to baseline, it can take up to 16 seconds for the BOLD signal to return to baseline. It is therefore difficult to distinguish the BOLD signal for stimuli that are presented closely together. We therefore used condition blocks of 16 seconds for our block-design experiments (Chapters 5, 6). For our event-related experiments (Chapters 3, 4), trials were presented in a random, counterbalanced order so trials were preceded by all conditions equally. In these cases, each trial lasted 3 seconds and several trials were grouped together before classification took place.

2.6 fMRI data analysis

We used BrainVoyager QX (Brain Innovation, Maastricht, the Netherlands) to perform preprocessing of functional runs. We then performed a general linear model (GLM) on each run and also across all runs. We used custom-built multi-voxel pattern analysis (MVPA) software to classify functional data throughout all experiments. Additional statistical procedures (e.g., two-way ANOVA's) were conducted in SPSS and all figures were edited for publication using Adobe Illustrator.

2.6.1 Preprocessing of functional data

Necessary preprocessing was performed on both anatomical and functional MRI data. Slice scan time correction was applied to account for the delay in the peak of the hemodynamic response function. Temporal high-pass filtering was applied to remove low-frequency drifts

caused by the inhomogeneity of the scanner signal over time. Data was transformed into Talairach space so to aid in localisation of visual areas and to allow comparisons across participants where necessary (e.g., the between-subjects surface map of a searchlight classification).

2.6.2 General linear model

We computed the GLM on each voxel to explain the variation of the BOLD signal through a linear combination of several predictors (i.e., stimulus conditions). This requires a design matrix which specifies all events which were active at any point during the time course. Each predictor is fit with the HRF (see *2.5.1 The BOLD signal*) to create a model response for the time course of each predictor. Each predictor is given a beta-weight to quantify the potential contribution in explaining the BOLD response for that voxel, minimising the error. It can be written as:

$$\sum_{i=1}^{t} y_i = b_{i0} + b_{i1}X_{i1} + b_{i2}X_{i2} + \ldots + b_{ip}X_{ip} + e_i$$

where y is the BOLD signal response, X is a predictor, b is the beta-weight of that predictor, b_0 is the constant and represents the baseline level, e is the error due to noise fluctuations in the scanner, p is the number of predictors and t is the number of time points to be measured. This is repeated univariately for all measured voxels. Once we obtained the beta-weights, we then applied contrasts to perform statistical tests between our experimental conditions.

2.6.3 Multi-voxel pattern analysis classifier

The main statistical measure we used in the analysis of these experiments was how accurately an MVPA classifier (Cox & Savoy, 2003; Haynes & Rees, 2006) could discriminate between different stimulus conditions based solely on fMRI activity. The classifier uses small biases that are present in each voxel in the neural signal between different stimulus categories (**Figure 2.4a**) and examines the spatial pattern of these biases during each stimulus presentation (**Figure 2.4b**). In comparison, conventional univariate approaches examine the change in activity between stimulus conditions at thousands of locations individually and are then averaged across the spatial domain. Since the multivariate approach uses the relationship across voxels, sensitivity is dramatically increased and we are able to discriminate between stimulus categories that have underlying distributions that are not spatially distinct. This increase in sensitivity meant far fewer presentations were necessary in order to discriminate between stimulus conditions. Indeed, while a univariate approach usually requires minutes of stimulus presentation to identify a significant change of fMRI activity between conditions, this can be observed for single blocks and even single trials when using an MVPA classifier (LaConte, 2011). Nonetheless, additional stimulus presentations are required to generate training data which the classifier uses as the basis of its discrimination (the precise number depends on the experimental design).

However, there are potential disadvantages involved with using MVPA classification techniques. The informative voxels used for discrimination by the MVPA classifier are distributed in a way that is unnatural to the visual system and does not represent the anatomical connections of the human brain. There is also no guarantee that classification performance reflects a change of perceptual state. Stimulus categories may differ in ways that are undetectable to observers but which nevertheless create spatial patterns in the fMRI

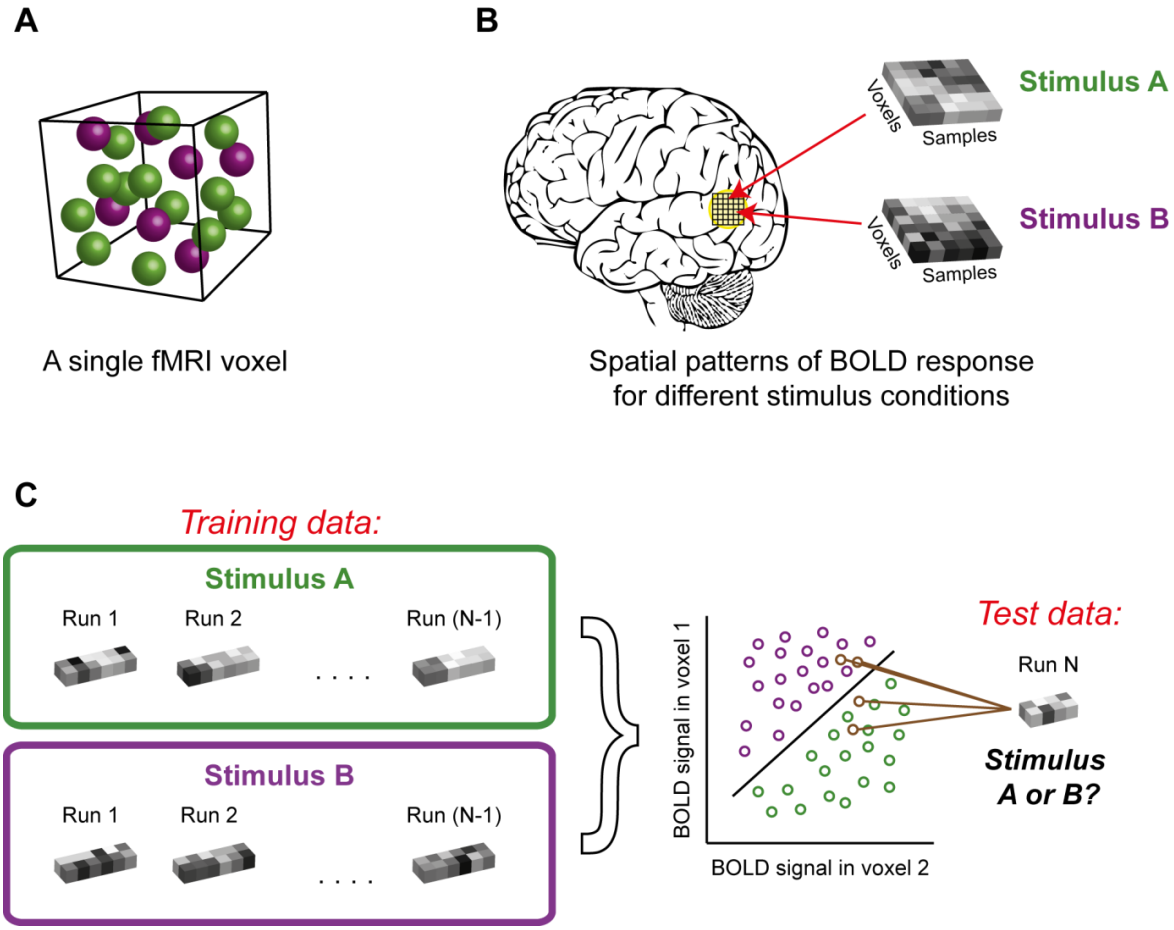


Figure 2.4: An outline of the MVPA procedure. (A) A single fMRI voxel which contains a small biases towards stimulus A (green spheres). (B) Distinct spatial patterns are observed for different stimuli which can then be discriminated by the MVPA classifier. (C) A sample classification using only two voxels. The BOLD responses for these two voxels are obtained across samples and runs and given a label depending on the stimulus category. These are plotted against one another and a hyperplane was drawn to best separate these categories. The classifier is then presented with independent test data and attempts to categorise the fMRI activity based on the model created from the training data.

activity that allow for successful decoding by the classifier. Further, the single algorithm used by the MVPA classifier may bias successful decoding of stimulus categories in certain visual areas. As such, classification performance may differ between visual areas due to changes in the underlying spatial distribution of the fMRI activity, rather than the ability of the classifier to discriminate between the stimulus categories.

For our experiments, we used a custom-built MVPA classifier that used SVM^{light} toolbox libraries (http://svmlight.joachims.org), and was performed using MATLAB 2009a.

Spatial patterns were generated using a subset of voxels within each visual area that were most responsive to the contrast "all stimuli versus fixation". This contrast was chosen to restrict voxel selection to eccentricities presented in the stimulus and increased the likelihood of the voxels being relevant to the processing of the stimulus. The patterns were used to create training data which represented the neural activity of each stimulus category. A support vector machine (SVM) algorithm was used to partition the training data, drawing a hyperplane which maximised the separation between the categories. An example of this partition for two voxels is presented in **Figure 2.4c**. This was typically repeated for a large number of dimensions (i.e., voxels) such that we could learn a general mapping between patterns of fMRI activity and stimulus conditions.

The classifier was then presented with independent test data and asked to identify the stimulus category this activity represents based on the model created from the training data (**Figure 2.4c**). We used a mean-leave-one-run-out cross-validation procedure, meaning training data was generated for all runs except for one run which was used as the test data. This was alternated such that each run was used as test data once, with the cross-validation reducing variability in the results while ensuring there was no dependence between training and testing data. The prediction accuracy of each validation was combined to give the mean prediction accuracy; a value between 0 – 1 indicating how often the classifier was able to correctly identify stimulus category.

We generally selected 300 voxels for use in the classification patterns for each visual area. This number was chosen as it was towards the upper bound of voxels available in smaller visual areas (e.g., V3v), so that we did not bias voxel selection towards visual areas that contained more voxels and could therefore generate larger patterns. We plotted prediction accuracies as a function of pattern size to ensure prediction accuracies had become saturated

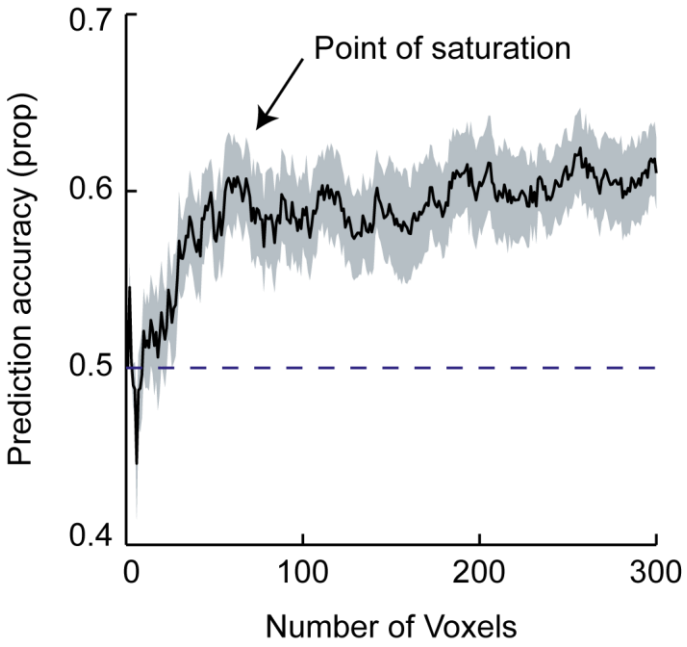


Figure 2.5: Sample data showing prediction accuracies as a function of pattern size. The arrow indicates the number of voxels where prediction accuracies saturated, and increasing pattern size beyond this had minimal effect on the performance of the classifier. Shaded region is the standard error. The dotted line is chance performance for the binary classifier. This example data was the between-subjects mean in VIPS for the signal-in-noise task (Chapter 3).

with this number of voxels (e.g., **Figure 2.5**) and that we were not biasing our results through this arbitrary decision. Further, we excluded ROIs in participants where less than 75 voxels were available after the voxel selection contrast was applied. This was due to the unstable nature of classification accuracies before saturation took place (e.g., **Figure 2.5**).

3. Signal-in-noise and feature difference tasks

Abstract

The cortical processing of coarse and fine binocular disparities is suggested to be specialised in the dorsal and ventral pathways respectively. However, previous examinations of this proposal have consistently measured coarse disparities using a signal-in-noise task and fine disparities through the discrimination of feature differences. We decoupled these properties and investigated the influence of the perceptual task when these were presented within the same range of disparities. We used human fMRI to examine the relationship between cortical activity and the perception of depth in both signal-in-noise and feature discrimination tasks. For the signal-in-noise task, we varied the proportion of dots located at the depth of the target plane. In the feature difference task, we varied the relative disparity between the target plane and its surround. We manipulated participants' performance on these tasks and recorded event-related fMRI responses obtained concurrently. We used a multi-voxel pattern classifier to identify cortical regions that varied in line with perceptual judgments. In both tasks we indicated a close association between psychophysical judgments of depth and activity in parietal regions VIPS and POIPS. We suggest these regions integrate depth information and form the percept of a 3D surface that is used in later parietal regions to carry out visuomotor tasks such as reaching or grasping.

3.1 Introduction

The binocular disparities that result from having two horizontally separated eyes provide a powerful cue to the three-dimensional (3D) structure of the surrounding environment. This information is potentially useful in supporting a range of tasks that depend on inferring the distance of objects [for example, from grasping an object (Culham et al., 2003; Sakata, Taira, Kusunoki, Murata, & Tanaka, 1997; Sakata et al., 1998; Theys, Pani, van Loon, Goffin, & Janssen, 2012a) to recognising complex aspects of the 3D shape (Chandrasekaran et al., 2007; Durand et al., 2007; Durand et al., 2009; Georgieva et al., 2009; Orban, 2011; Srivastava et al., 2009; Theys et al., 2012b)]. Performance on such tasks is likely to rely in part on disparity representations that have been observed throughout the visual, temporal and parietal cortices (Gonzalez & Perez, 1998; Orban et al., 2006b; Parker, 2007). However, the mapping between responses in different cortical areas map and the execution of different tasks is uncertain.

One means of conceptualising the processing of disparity information in the primate visual system is that there are specialisations in the dorsal and ventral processing streams. In particular, dorsal regions are suggested to process coarse disparity signals while ventral regions process fine disparities (Neri, 2005; Tyler, 1990). Evidence compatible with this suggestion comes from a number of neurophysiological studies that have measured single unit responses while macaques were engaged in tasks that involved either the discrimination of a target depth plane embedded within a noisy background (‘coarse’ task) or the discrimination of small differences in depth (‘fine’ task). Specifically, responses of individual neurons in dorsal region MT (V5) have generally been found to be as sensitive as that of the monkey when they were engaged in a coarse task (Uka & DeAngelis, 2003, 2004). Further, electrical microstimulation of MT neurons was found to bias perceptual judgments on the coarse task (DeAngelis et al., 1998; Uka & DeAngelis, 2006), but not for the fine task (Uka &

DeAngelis, 2006) thereby providing evidence of a dissociation based on disparity magnitude in the dorsal visual pathway. In contrast, performance on the fine task was affected by microstimulation in ventral region V4 (Shiozaki et al., 2012) and neurons at the end stages of the ventral pathway have been found to respond to fine changes in depth position (Janssen et al., 2000b; Uka et al., 2005).

While these results are consistent with a separation between pathways, the characterisation of responses across the whole pathways is difficult based on samples from a small number of areas. Recent work using brain imaging approaches has tested cortical responses to a range of disparity magnitudes (from fine to coarse). This work suggests significant modulation of activity in the dorsal visual pathway (occipital and parietal regions) to both fine and coarse scale disparity signals (Backus et al., 2001; Cottereau et al., 2011, 2012a; Minini et al., 2010; Preston et al., 2008). This suggests that any separation between pathways in terms of the magnitude of disparity is not very clear cut.

Here rather than considering disparity magnitude *per se*, we set out to test the potential involvement of different cortical areas in supporting performance in 'coarse' (signal-in-noise) and 'fine' (feature difference) tasks that have previously been used extensively. To this end, we rendered depth planes in random-dot stereograms (RDS) and sought to manipulate participants' behavioural ability to judge depth position (near *vs.* far) in one of two ways (**Figure 3.1a–b**). In Experiment 1, we varied the percentage of dots defining the position of the target plane in relation to dots with randomly assigned disparities (i.e., we varied the signal-to-noise ratio using external noise manipulations, a 'coarse' task). In Experiment 2, we titrated the disparity difference between a target plane and its surround (i.e., a 'fine' task that is limited by internal noise). We considered fMRI responses in pre-defined regions of interest in the visual and parietal cortices, and use multi-voxel pattern analysis (MVPA) to determine

the extent to which near *vs.* far depth positions could be decoded based on the fMRI activity. We assessed changes in classification performance and behavioural performance as the signal strength was manipulated, to test for areas whose activity changes in a manner consistent with the observer's perceptual judgments (Britten, Shadlen, Newsome, & Movshon, 1992; Chandrasekaran et al., 2007). In particular, we compared psychometric functions with the 'fMR-metric' functions generated using the MVPA classifier.

3.2 Experiment 1: Signal-in-noise task

To test the processing of disparities in a signal-in-noise task, participants discriminated the depth position (closer or farther than the fixation plane) of a target plane whose visibility was corrupted by a varying proportion of noise dots. This 'disparity noise' consisted of dot elements presented at random disparities chosen from a uniform distribution that straddled the fixation plane. This stimulus was modelled on that used in previous single-unit recording studies (DeAngelis et al., 1998; Uka & DeAngelis, 2003, 2004, 2006), although the range of disparities was reduced to make the comparison between tasks unrelated to gross differences in disparity magnitude. In addition, to decouple the perceptual interpretation of the stimulus from the overt behavioural response (a button press) used by participants to report their perception we used a delayed cue response paradigm (Zhang, Meeson, Welchman, & Kourtzi, 2010). In particular, a cue was presented on each trial (a red square or a green triangle) that changed the key-press mapping used to indicate near *vs.* far such that a given motor response was uncoupled from the depth interpretation.

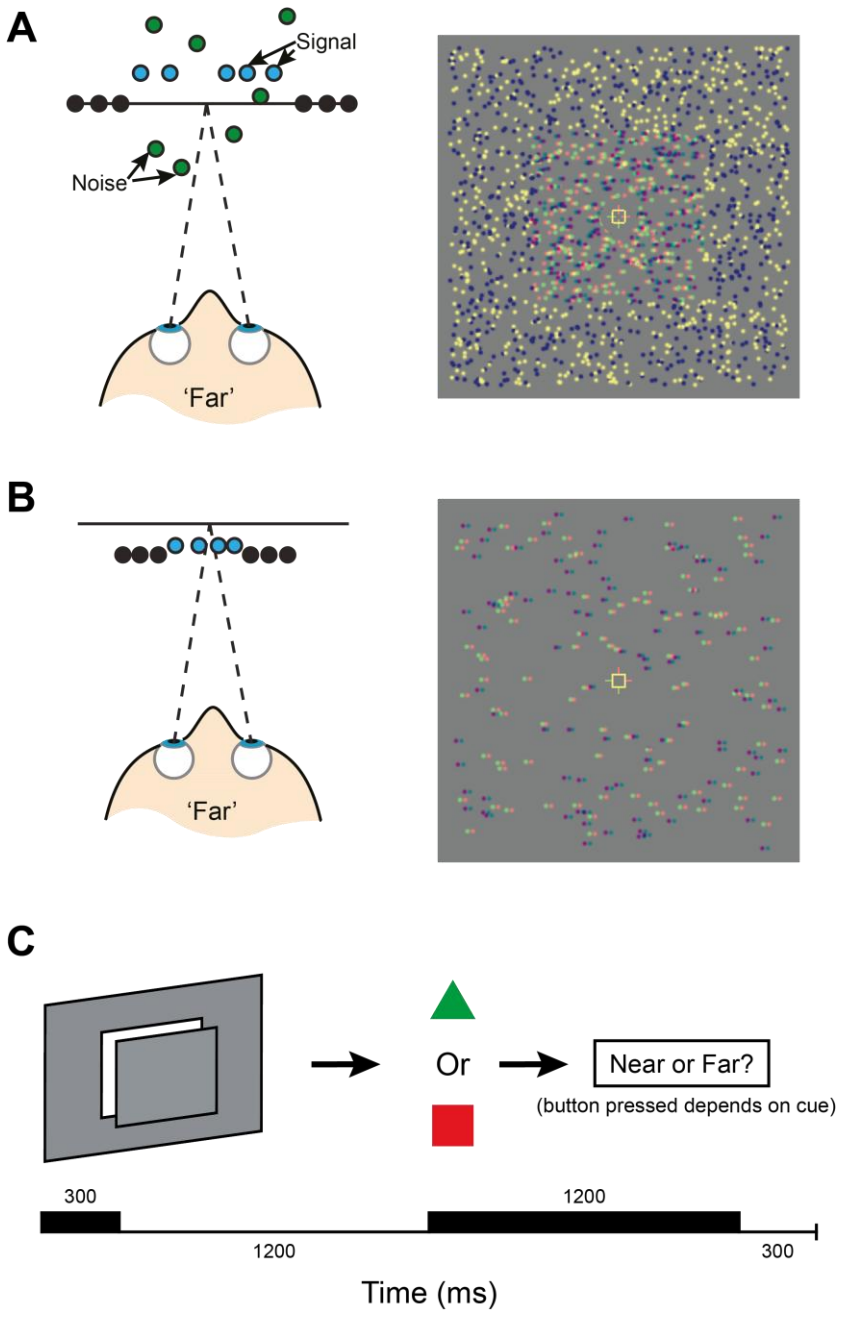


Figure 3.1: Stimuli and design of signal-in-noise and feature difference tasks. (A) A representation of stimuli in the signal-in-noise task viewed from above, and an illustrative red-green stereogram of the stimulus. Only a certain percentage of dots are located at the correct depth plane; the remainder are given random disparities. In the left image, blue dots represent 'signal' dots that were located at the target disparity, green dots represent distracting 'noise' dots, and black dots represent the surrounding pedestal disparity, located at the plane of fixation (horizontal line). In the experiment, all dots were black and white. (B) A representation of stimuli in the feature discrimination task viewed from above, and an illustrative stereogram of the stimulus. The pedestal was always located at a crossed disparity and the central test plane was titrated around this. Although all dots are given a crossed disparity, there is a central patch of dots that have a smaller disparity and appear behind the surrounding dots; therefore this stimulus is classified as 'far'. (C) The procedure for a typical trial. The first black bar indicates stimulus onset and duration, the second black bar indicates presentation of the response cue.

3.3 Methods: Signal-in-noise task

3.3.1 Participants

Eleven participants (four females) were recruited from the University of Birmingham, including one of the authors. Mean age was 25.3 years (range, 19 – 34 years). All participants had normal or corrected-to-normal vision with no deficits in colour vision and were paid for their participation. Prior to scanning, all participants were screened for stereo deficits in the laboratory to ensure they could discriminate at least 1 arcmin of disparity. Two further participants were tested, but subsequently excluded from further data analysis. The first was excluded due to excessive head movement during scanning (their head moved more than 7 mm or 7 degrees drift over the course of the scan) as this disrupts voxel co-registration and is thus incompatible with the logic of MVPA decoding. The second participant was excluded due to poor behavioural performance in the scanner (their discrimination performance was <56% for the 60% signal condition and below). The experiment was granted ethics approval by the University of Birmingham STEM ethical review committee and all participants provided written informed consent.

3.3.2 Stimuli

We used random-dot stereograms (RDS) defined by black and white dots within a rectangular aperture (14 × 19°) and displayed on a mid-grey background. Within this region, a central test plane (7 × 7°) was given a nonzero disparity of ±6 arcmin relative to the fixation plane. To minimise the effects of adaptation, some jitter (up to ±1 arcmin) was added to this disparity value for each trial. The dot density of the stereogram was 8 dots/deg² and each dot had its edges softened using a Gaussian distribution (diameter 0.15° at half-height). The background surrounding the RDS was filled with a grid of black and white squares which could be used as

a clear reference and encouraged vergence to the plane of the screen. The fixation marker was a hollow white square (0.5° side length) that was presented in the centre of the stereogram. Vertical and horizontal nonius lines (length 0.375°) were attached to the fixation marker to promote vergence at the plane of the screen. We restricted presentation of dots to outside of a circular region (1.5° diameter) centred at fixation to reduce interference from the stimulus on binocular fusion. We manipulated task difficulty by introducing noise dots that were located at random depths, up to a maximum of ±20 arcmin (**Figure 3.1a**). In the 0% signal condition, there is no consistent depth plane and all dots appear scattered in depth, while in the 100% signal condition no noise dots were presented. We employed five different signal levels (0, 20, 40, 60 and 100%) to sample different levels of psychophysical performance. Under all conditions, dots outside the target were located at the plane of the screen and contained no disparity noise.

3.3.3 Design

For each participant, we collected both behavioural and fMRI data concurrently in a single session of approximately nine event-related scans. On each trial, participants made a perceptual judgment on the depth sign of the stimulus (i.e., 'near' or 'far'). Trials were randomised and counterbalanced such that each condition preceded all other conditions equally (for one trial back), with the order of trials differing between runs and participants. Eleven conditions were presented on each run: 10 stimulus conditions (5 signal levels × crossed *vs.* uncrossed disparities) and one fixation condition during which the central test plane was removed. We collected 11 repetitions of each trial type on each run (total 121 trials) and added a single dummy first trial to ensure that trial history of the second trial was

balanced. Each scan started and ended with a 9 s fixation interval, and total duration of a single scan was 6 min 24 s.

Experimental trials lasted 3 s (**Figure 3.1c**) and started with a stimulus presentation of 300 ms followed by a delay of 1200 ms (75% of trials) or 1400 ms (25% of trials) during which the central test plane was removed from the screen. These different delay times were chosen to minimise predictability and habitual responding by participants. After the delay, a green triangle or red square appeared inside the fixation marker and served as an indicator for the motor response mapping to be used on that trial. In particular, if the response cue was a green triangle observers used a particular finger-key matching (e.g., index finger for 'near'), while if the response cue was a red square, observers reversed the finger-key matching (e.g., index finger for 'far'). This was balanced across participants to remove any bias for a particular cue. The response cue was removed 300 ms before the start of the next trial. This procedure separated the motor response (i.e., button press) from the perceptual depth judgment. Participants were familiarised with this process in the laboratory prior to scanning. During fixation trials, the fixation square and surround RDS were simply displayed for 3 s.

3.3.4 fMRI data acquisition

The study was performed in a 3-Tesla Philips Achieva MRI scanner at the Birmingham University Imaging Centre. We used an eight-channel SENSE head coil to collect echo-planar imaging (EPI) and T1-weighted anatomical (1 × 1 × 1 mm) data. For experimental runs, EPI data [echo time (TE), 35 ms; repetition time (TR), 1500 ms] were obtained from 25 slices (voxel size, 2 × 2 × 2 mm, 256 volumes) positioned close to the coronal plane. Localisers were obtained in a separate session, with EPI data (TE, 34 ms; TR, 2000 ms) acquired from 28 slices (voxel size, 1.5 × 1.5 × 2 mm, near coronal). We used a pair of video projectors

(JVC D-ILA SX21) to display stereoscopic images, with each projector fitted with a unique interference filter (INFITEC, GmBH) that distributed the wavelengths of visible light between the two projectors. The images from each projector were optically combined using a beam-splitter cube and passed through a waveguide into the scanner room. Stimuli were projected onto a translucent plastic screen located behind the head coil and inside the bore of the scanner. Participants viewed the screen via a mirror positioned on the head coil angled at 45°, with an optical path length of 65 cm. Since the INFITEC filters produced negligible overlap between the spectra for each projector, there was minimal crosstalk between the projected images for a participant wearing a pair of corresponding filters. Unique stimuli were pre-generated for each participant using C#, and the experiment was performed using MATLAB (The MathWorks, Natick, MA) and the PsychToolBox 3 extension (Brainard, 1997; Pelli, 1997).

For each participant, we identified regions of interest (ROIs) from independent data in a separate localiser scan. We used a rotating wedge stimulus and expanding concentric rings to define the borders of early retinotopic areas (V1, V2, V3v, V4, V3d, V3A and V7) (Aguirre, Zarahn, & D'Esposito, 1998; DeYoe et al., 1996; Sereno et al., 1995). In particular, V4 was defined as the ventral region adjacent and inferior to V3v containing a full hemifield representation (Tootell & Hadjikhani, 2001; Tyler et al., 2005) and V7 was defined as the dorsal region adjacent and anterior to V3A, also containing a full hemifield representation (Tootell et al., 1998; Tsao et al., 2003; Tyler et al., 2005). We identified higher dorsal regions [V3B/kinetic occipital area (KO), human motion complex (hMT+/V5)] and the ventral lateral occipital region (LO) from additional localiser scans. Area V3B/KO (Dupont et al., 1997; Zeki, Perry, & Bartels, 2003) was defined anatomically as the dorsal region adjacent and lateral to V3A with which it shared a foveal representation (Tyler et al., 2005), and

functionally as the set of voxels that responded significantly more ($p < .001$) to kinetic boundaries than to the transparent motion of black and white dots without clear borders. Area hMT+/V5 was defined as the region in the lateral temporal cortex that responded significantly more ($p < .001$) to an array of coherently moving dots than to an array of static dots (Zeki et al., 1991). The lateral occipital (LO) area was identified as the region in the lateral occipito-temporal cortex that responded significantly more ($p < .001$) to images of objects and shapes than to scrambled versions of these images (Kourtzi & Kanwisher, 2000, 2001). During all localiser scans, participants performed an attentionally demanding task on the fixation point, except for the LO localiser in which they had to respond if the same image was presented consecutively. For two subjects who also participated in the feature difference task, the ventral (VIPS) and parieto-occipital (POIPS) regions along the intraparietal sulcus were identified by the set of voxels that responded significantly more ($p < .001$) to 3D shape formed by disparity and structure-from-motion cues than random disparities and motion speeds which did not form 3D shape (Chandrasekaran et al., 2007; Orban, Sunaert, Todd, Van Hecke, & Marchal, 1999). For remaining participants, these regions were defined as regions anterior to V7 which showed significantly stronger ($p < .001$) responses to all of the experimental conditions in contrast to the fixation baseline. We were unable to obtain data for the parietal region POIPS for one participant due to the spatial resolution of the EPI sequence and near coronal slice positioning during the fMRI acquisition process. For illustrative purposes, the mapping of these ROIs for one participant is presented in **Figure 3.2**.

3.3.5 fMRI data analysis

We processed MRI data using BrainVoyager QX (BrainInnovation, Maastricht, the Netherlands). For each participant, anatomical scans were transformed into Talairach space

and used for 3D cortex reconstruction, inflation, flattening and the segmentation of gray and white matter. Preprocessing of functional data included head movement correction (translation and rotation), slice scan time correction, and removal of low-frequency drifts and linear trends through temporal high-pass filtering (3 cycles). Functional runs were aligned to the participant's anatomical data and then transformed into Talairach space. Consecutive volumes that contained head movement greater than 1 mm of translation or 1° of rotation were excluded from further analysis.

We used Multi-Voxel Pattern Analysis (MVPA) software to analyse the fMRI data from each ROI. For voxel selection, gray matter voxels were isolated through a grey-matter mask from the anatomical scan and only included voxels that showed a *t* value larger than 0 when contrasting fMRI activity from all stimulus conditions to the fixation baseline across all experimental runs. These were then sorted according to their response (*t* statistic) and the 300 most active voxels (or highest number available) in each ROI were used in the classification. Estimation of fMRI responses to single events in our event-related fMRI design were likely to be noisy for single trials; therefore, prior to feeding the data to the machine learning classifier we averaged a small number of trials (4, 4 and 3 from a single run) to generate 3 training patterns per run. Each voxel's time series was then normalised (*z*-score) in each experimental run to compensate for the decrease in mean image intensity with distance from the receiver coil. The fMRI time series was shifted by 3 volumes (4.5 s) to account for the hemodynamic delay of the BOLD signal and each volume had the mean univariate signal subtracted from it.

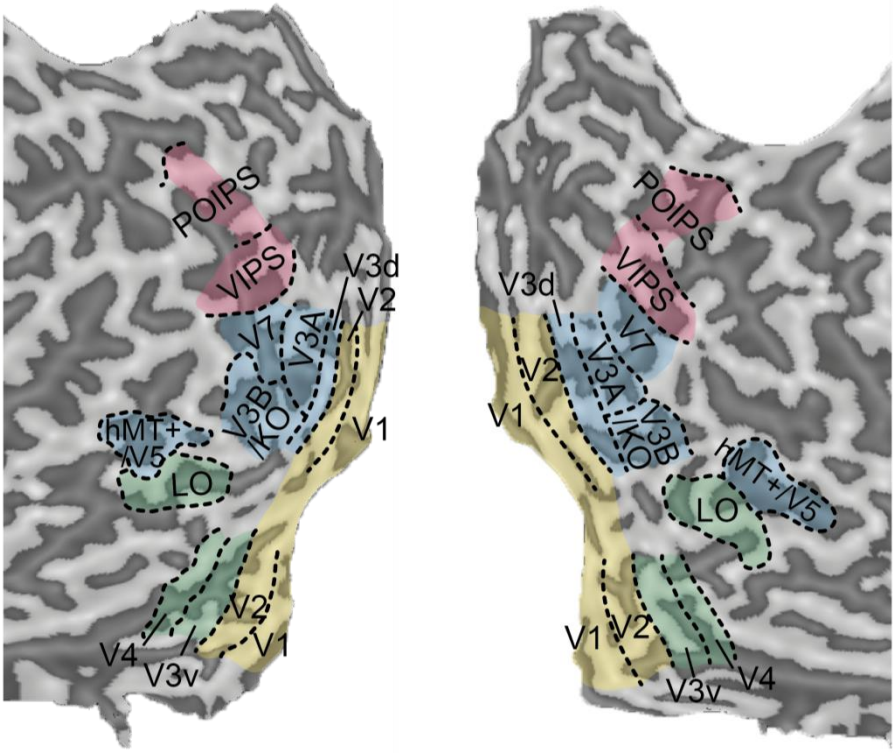


Figure 3.2: Regions of interest in one participant showing retinotopic areas, V3B/KO, hMT+/V5, LO and parietal regions VIPS and POIPS.

We used a leave-one-run-out cross-validation analysis procedure. For each cross-validation, one run was left out as an independent test dataset and the data from the rest of the runs was used as the training set. The classification accuracy for each ROI was obtained by averaging the prediction accuracy across cross-validations. Training was performed on the 100% signal condition, and test patterns were calculated for all five signal levels. The reported prediction accuracy of the classifier corresponds to the proportion of trials on which the algorithm could correctly predict depth sign based on the pattern of fMRI activity, in which chance performance would be 0.5 for a binary classification (i.e., 'near' *vs.* 'far' stimuli). For each ROI, prediction accuracies were averaged across participants. Examining

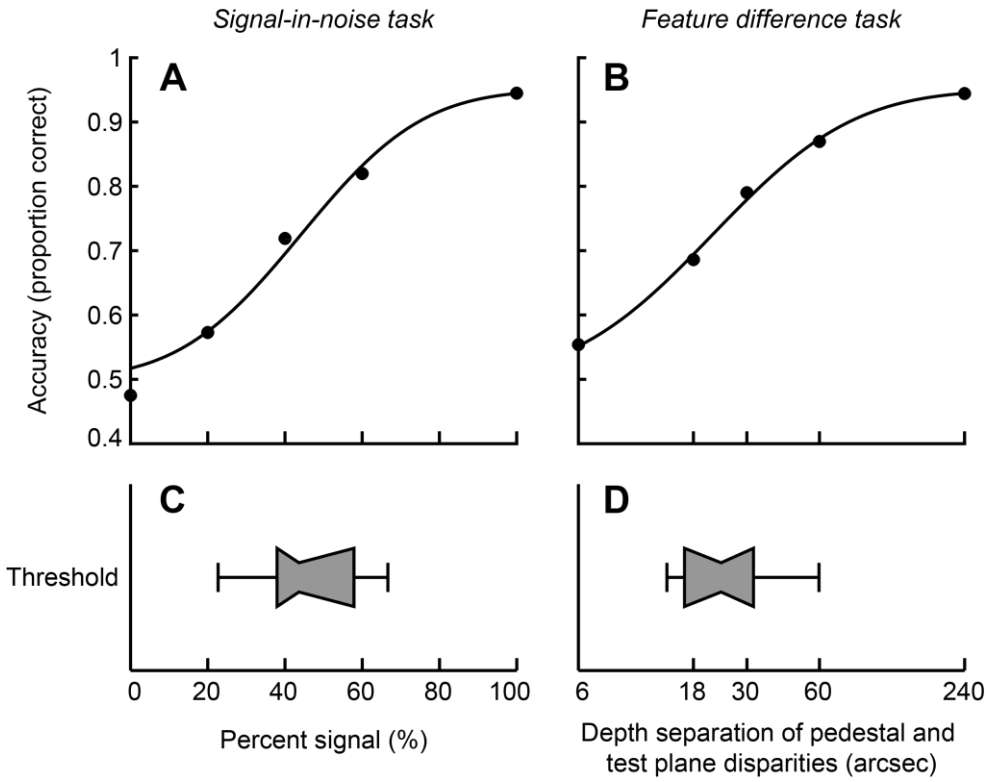


Figure 3.3: The mean behavioural results for both experiments. (A) Behavioural results for the signal-in-noise task, where the signal reflects the percentage of dots that were at the disparity of the test plane and accuracy refers to the proportion of responses that the participant correctly identified as near or far. The solid curve is the best-fitting Gaussian. (B) A similar function for the behavioural results for the feature difference task placed on a logarithmic scale. (C) Between-subjects threshold for the behavioural results for the signal-in-noise disparity task. The centre of the 'bowtie' represents the median, the shaded area depicts 68% confidence values, and the error bars are the 95% confidence intervals. (D) The corresponding graph for threshold values from the feature difference task.

prediction accuracies across pattern size (number of voxels) showed that classification values had saturated by 300 voxels in all of the ROIs we considered.

3.4 Results: Signal-in-noise task

We presented participants with random-dot stereograms that depicted a central target region that was either in front of or behind the surround background. We instructed participants to judge the depth position (near/far) of the central target while we measured fMRI responses in regions of interest in the visual and parietal cortices. We manipulated the difficulty of the

behavioural task by changing the proportion of signal dots in the central target relative to noise dots that had a randomly chosen disparity. As the proportion of noise dots increased—and fewer signal dots were present—the task became increasingly difficult. Based on pilot testing, we selected five different signal levels to use during fMRI scanning that sampled different locations on the psychometric function. To describe psychophysical performance, we fit the behavioural judgments with a cumulative Gaussian (**Figure 3.3a**) to estimate the 75% correct discrimination threshold (**Figure 3.3c**). On average, participants needed 44% signal dots to reliably judge the depth position of the target. This threshold is somewhat higher than those measured in macaque monkeys (typically between 10 and 20%: Uka & DeAngelis, 2003, 2004) however our stimuli had a smaller range of disparities and our presentation was much briefer (300 *vs.* 1500 ms).

We examined fMRI responses sampled from early visual areas (V1, V2), ventral regions (V3v, V4, LO), dorsal regions (V3d, V3A, V3B/KO, V7, hMT+/V5) and parietal regions (VIPS, POIPS). Using the measured fMRI responses, we trained a linear support vector machine to associate patterns of voxel activity within each ROI to the disparity-defined depth position of the stimulus that gave rise to the activity. We tested whether we could predict the viewed stimulus from the fMRI activity, calculating the mean leave-one-run-out prediction accuracy for classifiers trained to discriminate crossed from uncrossed disparities when no noise was present.

Figure 3.4a shows the between-subjects mean prediction accuracies obtained for the most discriminable stimulus configurations (100% signal) for each ROI. To establish a baseline for chance performance, and thereby judge responses that were statistically reliable, we ran the classification analysis with randomly permuted fMRI patterns (i.e., we randomised the correspondence between fMRI data and training labels and estimated the classifier

prediction for each visual area) over 999 iterations for the 100% signal condition. This created a distribution of classification accuracies, and we used the upper 99.5% centile (one-tailed, Bonferroni corrected) as our criterion for statistical significance (**Figure 3.4**, dotted lines). For all regions of interest, the median of the shuffled distribution was very close to 0.5 (range, 0.498 – 0.501) confirming our analysis technique to be unbiased. With the exception of hMT+/V5, all measured dorsal and parietal visual areas supported classification accuracies that exceeded the criterion for chance decoding accuracies. In contrast, only region LO of the ventral stream had responses that were comfortably higher than chance, while earlier ventral regions (V3v, V4) remained very close to this threshold.

To test whether there was a relation between behavioural judgments of depth position and information about depth position contained in the fMRI responses, we evaluated the decoding performance of the classifier at different signal levels. Thereby, we generated an 'fMR-metric' function, which we then fit using a cumulative Gaussian (**Figure 3.5**, dashed lines). To compare our ability to decode depth positions with the behavioural performance of our human participants, we used the parameters (mean, threshold) of the psychometric function (**Figure 3.3a**) to constrain the Gaussian fit to the fMRI data, allowing only the maximum and minimum values to vary as free parameters. This created a scaled version of the behavioural results and allowed us to make direct comparisons between the simultaneously recorded fMRI activity and behavioural performance. We present these fMR-metric functions for ROIs with above chance performance in **Figure 3.5** and performed a χ^2 goodness-of-fit test to quantify the fit of the fMRI classification accuracies to the values on the scaled fMR-metric function (**Table 3.1**). fMRI responses in both early (V3d, V3A) and later (VIPS, POIPS) dorsal visual areas could be decoded in a manner that was similar to the behavioural performance. We also observed a marginal result in V3B/KO ($r = .873$, $p = .053$)

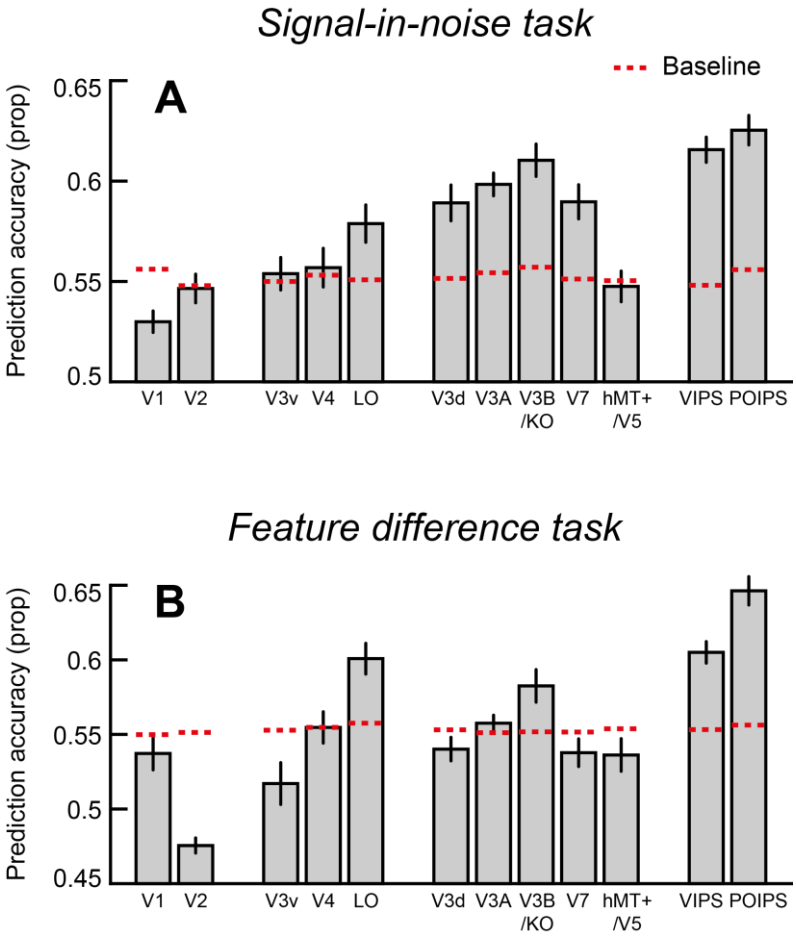


Figure 3.4: Prediction accuracies for each visual area. (A) The mean prediction accuracy of the classifier for the 100% signal condition in the signal-in-noise task. The horizontal red lines mark the baseline of statistical significance generated from permuting the data labels before being fed into the classifier. Error bars are standard error. (B) The mean prediction accuracy of the classifier for the 240 arcsec condition in the feature difference task.

and V7 ($r = .838$, $p = .076$); however best performance was observed for fMRI responses in the parietal cortex. In contrast, we did not observe a similar parametric effect in ventral region LO ($r = .690$, $p = .197$), with performance seen to deteriorate as soon as noise was introduced into the stimulus.

These results suggest a similarity between perceptual judgments of depth position in a signal-in-noise task and fMRI responses in higher portions of the dorsal visual stream. In contrast, introducing external noise to the display appears to severely disrupt our ability to decode near *vs.* far depth positions in higher ventral region LO. These findings for the signal-

in-noise task are consistent with the suggested involvement of the dorsal visual pathway in processing signals relevant to a coarse disparity task (DeAngelis et al., 1998; Uka & DeAngelis, 2003, 2004, 2006). We found the clearest association between fMRI responses and perceptual responses in parietal regions VIPS and POIPS. While homologues of these areas have not been studied extensively for coarse disparity tasks in the macaque, our findings are consistent with recent neuroimaging evidence. In particular, parietal areas support preferential decoding of correlated stereograms (which support depth percepts) over anticorrelated stereograms (that do not support a perceptual interpretation of depth position) (Preston et al., 2008) and these regions respond differentially to the disparity magnitude of a sinusoid in a manner broadly consistent the perceptual interpretation (Minini et al., 2010).

3.5 Experiment 2: Feature difference task

The results of Experiment 1 are broadly consistent with the notion that dorsal visual areas are involved in the processing of perceptually-relevant signals for 'coarse' disparity tasks that involve separating a signal from external noise (Chowdhury & DeAngelis, 2008; Shiozaki et al., 2012; Uka & DeAngelis, 2006). Moreover, we observed that we could decode depth positions from ventral area LO under 100% signal conditions, but performance deteriorated rapidly once a small amount of noise was introduced. In Experiment 2, we sought to consider the roles of visual areas in the processing of fine disparity differences that are limited by internal processing noise (rather than external perturbations of the displays introduced by the experimenter). To this end, we asked participants to judge small differences in the relative disparity between a target plane and its surround (which was located in front of the fixation plane). For this stimulus, we expected that task performance would be limited by internal

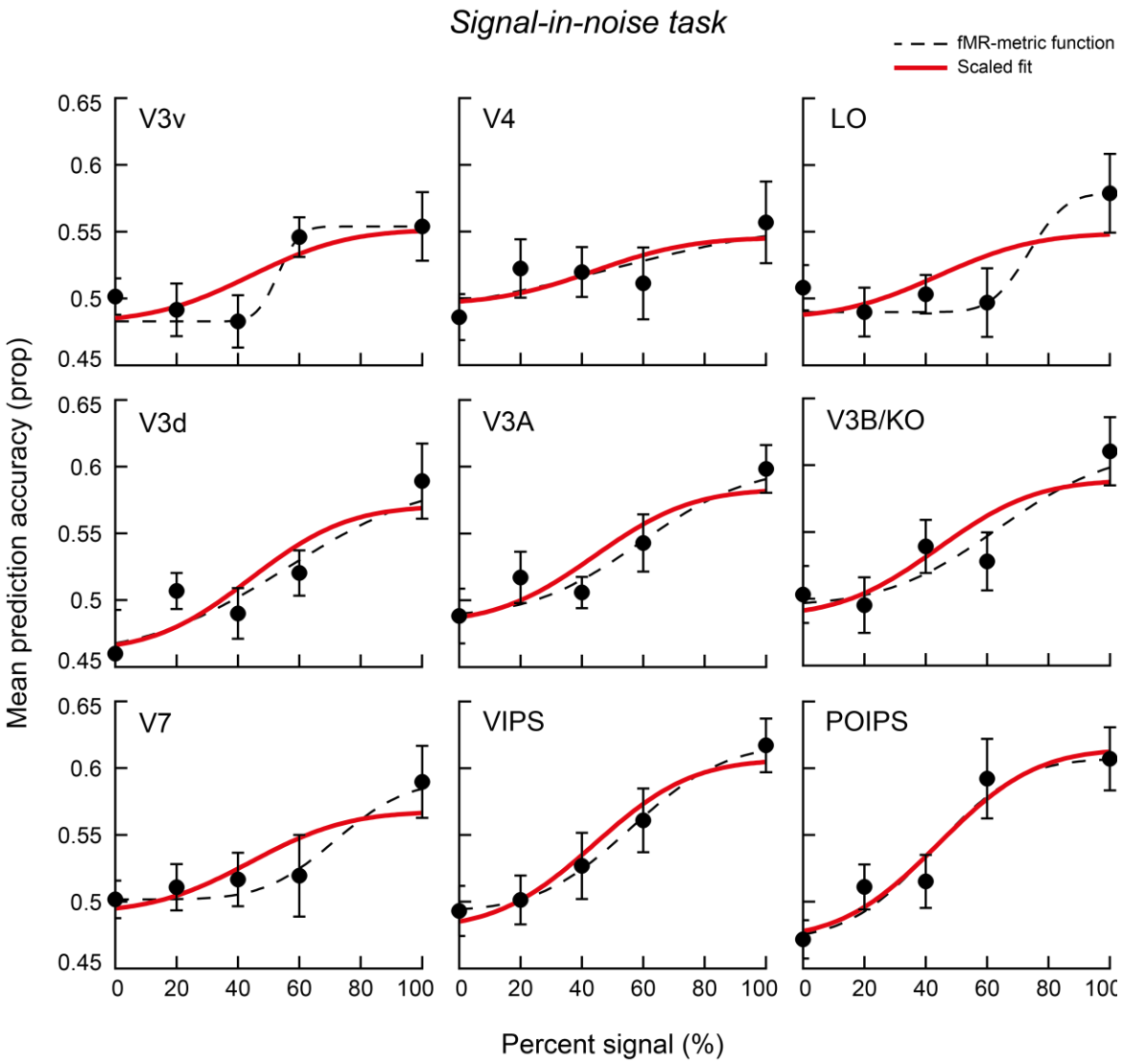


Figure 3.5: Mean fMR-metric functions for the signal-in-noise disparity task. The red line is a scaled version of the behavioural results and the dotted line is the best-fitting Gaussian. We have presented only visual areas that performed significantly above chance. Error bars are standard error.

noise, thus all of the dots in the stimulus carried signal (i.e., there was no external noise as there was in Exp. 1).

3.6 Methods: Feature difference task

Eleven participants (five females) were recruited from the University of Birmingham, including both authors. Mean age was 28.7 years (range, 20 – 35 years). One further

	Pearson Corr	P-value
V3v	.831	.082
V4	.766	.131
LO	.690	.197
V3d	.885	**.046***
V3A	.914	**.030***
V3B/KO	.873	.053
V7	.838	.076
VIPS	.978	**.004***
POIPS	.967	**.007***

Table 3.1: Goodness-of-fit of the fMR-metric functions for the signal-in-noise task. A χ^2 goodness-of-fit test of the fMRI data points for the signal-in-noise disparity task (Figure 3.5) to the scaled version of the behavioural results (Figure 3.3a). An asterisk marks a significant fit for that ROI ($p < .05$).

participant was excluded from the data analysis due to poor behavioural performance (accuracy below 57% across all conditions) and another three were excluded for excessive head movement (greater than 7 mm or 7 degrees drift over the course of the scan). Except for the differences outlined below, the methods were the same as for Experiment 1.

We presented RDS consisting of two concentric squares (side lengths 7° and 14°) at different depths. Task difficulty was manipulated by titrating the disparity of the target with respect to the surrounding pedestal (**Figure 3.1b**). The pedestal had a crossed disparity of 12 arcmin and the target plane varied ±6, 18, 30, 60 or 240 arcsec around this. As before, we trained the classifier to discriminate near *vs.* far relative disparities on the most discriminable conditions (±240 arcsec disparity), and then evaluated the performance of the classifier at all five relative disparity levels. Given the high sensitivity of human vision to disparity signals (McKee, 1983; Westheimer & McKee, 1977), we used a sparse dot density of 1 dot/deg^2 to ensure the task remained difficult and we excluded dots in the neighbourhood of the fixation marker (circular exclusion zone 2° in diameter) to minimise interference with binocular fusion. Since both planes had nonzero disparities, we removed the entire RDS for the fixation condition, displaying only the fixation marker and background.

Localisers were performed as described for *Experiment 1*, with the addition that we used independent localisers for parietal regions VIPS and POIPS for all participants. This contrasted the activity from presentation of 3D shapes defined by disparity and structure-from-motion to random patterns of its components (Chandrasekaran et al., 2007). We were not able to obtain data from POIPS in one participant due the slice positioning necessitated by their neuroanatomy.

3.7 Results: Feature difference task

Participants were asked to decide whether the central presented plane was in front or behind the surrounding pedestal reference plane. We considered five different disparity separations, selected so as to characterise variations in psychophysical performance similar to those measured in Experiment 1. We fit the behavioural results using a cumulative Gaussian (**Figure 3.3b**) and estimated the discrimination threshold. We found that participants required around 21 arcsec disparity difference for the depth to be reliably perceived (75% threshold; **Figure 3.3d**) which is within the range of values observed for macaque monkeys performing a similar task (Shiozaki et al., 2012; Uka & DeAngelis, 2006).

Following the analysis approach used in Experiment 1, we first tested for areas in which we could reliably decode relative depth positions for the highest signal levels. We found above chance decoding performance in ventral areas V4 and LO, and dorsal areas V3A, V3B/KO, VIPS and POIPS (**Figure 3.4b**). Thereafter, we examined the performance of the classifier as the disparity difference between the target and its surround was varied, and computed fMR-metric functions by fitting a cumulative Gaussian sigmoid to the fMRI decoding data (**Figure 3.6**, dotted lines). In addition, we used the psychometric function to fit the fMRI data, allowing the maximum and minimum to vary as free parameters while

restricting the shape of the function to that of the behavioural results. We observed significant fits of the behavioural results to the fMRI data in dorsal (V3A, V3B/KO) and parietal (VIPS, POIPS) regions, while a marginal fit was observed in the ventral stream for LO (r = .840, p =

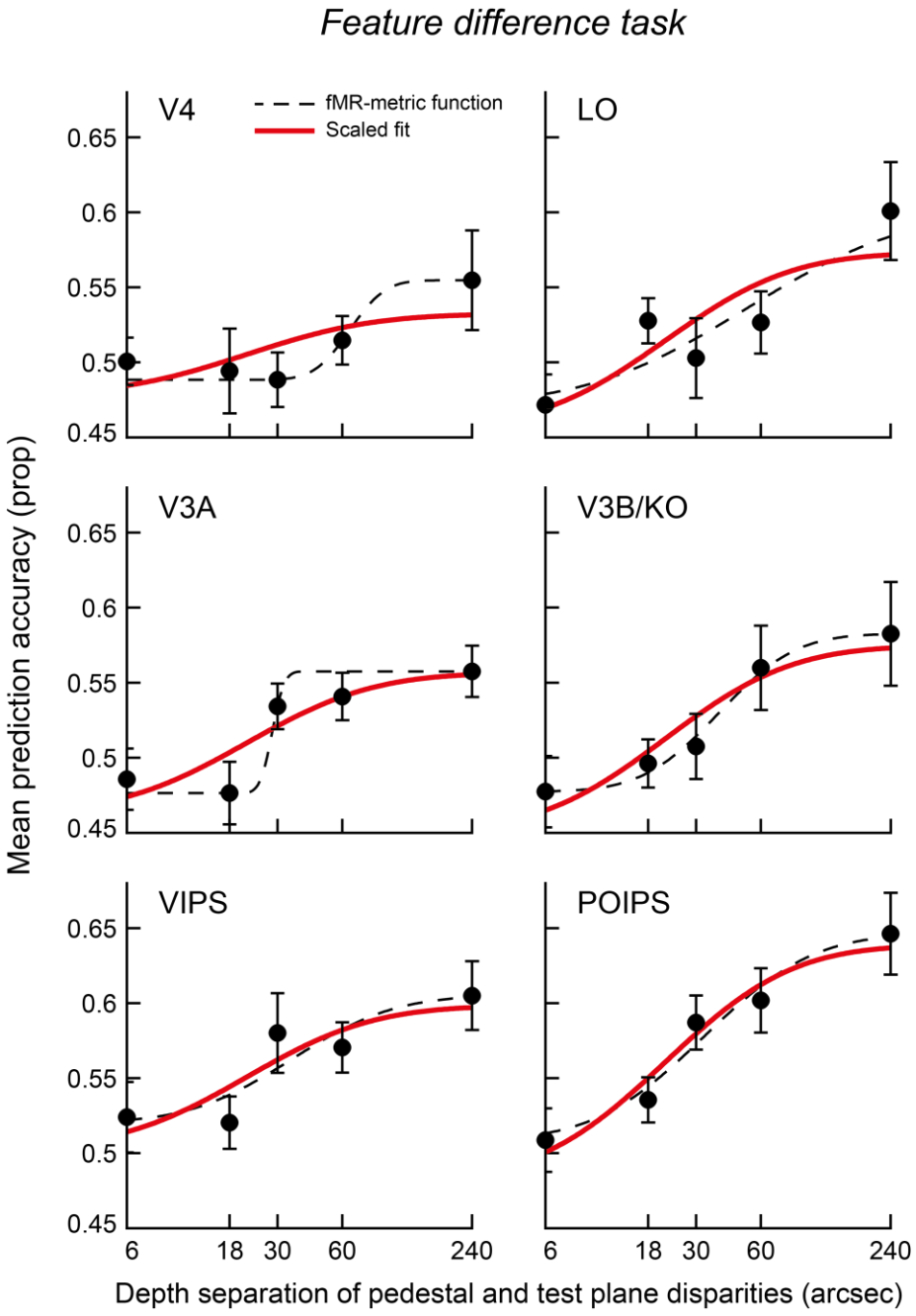


Figure 3.6: Mean fMR-metric functions for the feature difference task. The red line is a scaled version of the behavioural results and the dotted line is the best-fitting Gaussian. We have presented only visual areas that performed significantly above chance. Error bars are standard error.

	Pearson Corr	P-value
V4	.693	.194
LO	.840	.075
V3A	.892	**.042***
V3B/KO	.951	**.013***
VIPS	.883	**.047***
POIPS	.977	**.004***

Table 3.2: Goodness-of-fit of the fMR-metric functions for the feature difference task. A χ^2 goodness-of-fit test of the fMRI data points for the feature difference task (Figure 3.6) to the scaled version of the behavioural results (Figure 3.3b). An asterisk marks a significant fit for that ROI ($p < .05$).

.075) (**Table 3.2**).

Activity in parietal regions VIPS and POIPS showed the clearest relationship to depth judgments for the feature difference task and replicated our finding from the signal-in-noise task. However, these results were inconsistent with a neurophysiology study that found microstimulation of dorsal MT neurons did not bias behavioural performance when monkeys were engaged in a fine disparity task (Uka & DeAngelis, 2006), though this may simply relate to processing at a different stage of the dorsal pathway. Responses in other dorsal areas (V3A, V3B/KO) and ventral area LO were also reasonably well described by the behavioural model. However, the limited prediction accuracies supported by activity in these areas make it difficult to have high certainty about the contributions of these areas. In particular, lower levels of performance were close to chance levels, meaning that our detection power was limited

3.8 Control analyses

We performed several control measures and designed our stimulus to ensure that the results were best explained through changes in the perceptual depth interpretation. First, we recorded horizontal eye-movements for participants in both experiments using a monocular limbus eye-

tracker (CRS ltd). Analyses of these data suggested no systematic difference in eye-position (**Figure 3.7a–d**) and no statistical difference in the number of saccades between conditions, suggesting it was unlikely that eye-movements could adequately explain our results. The confines of the fMRI scanner and our use of spectral filters for stereoscopic presentation meant that we were unable to record changes in eye vergence; however we designed our stimuli to minimise this possibility. In particular, we instructed participants to maintain alignment of vertical and horizontal nonius lines that surrounded the fixation marker throughout all runs, encouraging vergence at the correct depth (the plane of the screen). Further, the stereogram was surrounded by a clear reference pattern that was located at the fixation plane. For the signal-in-noise task, the central test plane was also surrounded by a zero-disparity RDS present throughout all trials. However, for the feature difference task the pedestal RDS was presented at a crossed disparity on all trials and it is possible that participants' vergence state changed and became biased away from the fixation plane. To address this, before scanning commenced all participants were presented with each stimulus condition randomly for a total of 200 trials while participants undertook a vernier task (Popple et al., 1998). There were minimal differences between experimental conditions for thresholds of psychophysical performance and these were centred close to a vernier displacement of zero (**Figure 3.7e**). This suggests that observers were able to maintain stable vergence across conditions (perhaps assisted by the background reference marks and the fixation point) despite the vergence demand of the pedestal and test planes in front of the point of fixation.

Second, to ensure our results were related to disparity processing and not an overall change in the univariate signal, we computed the percent signal change (PSC) by comparing the difference in the BOLD signal between the highest signal level and fixation in each ROI. Similarly, we examined the functional signal-to-noise ratio (fSNR) by comparing the

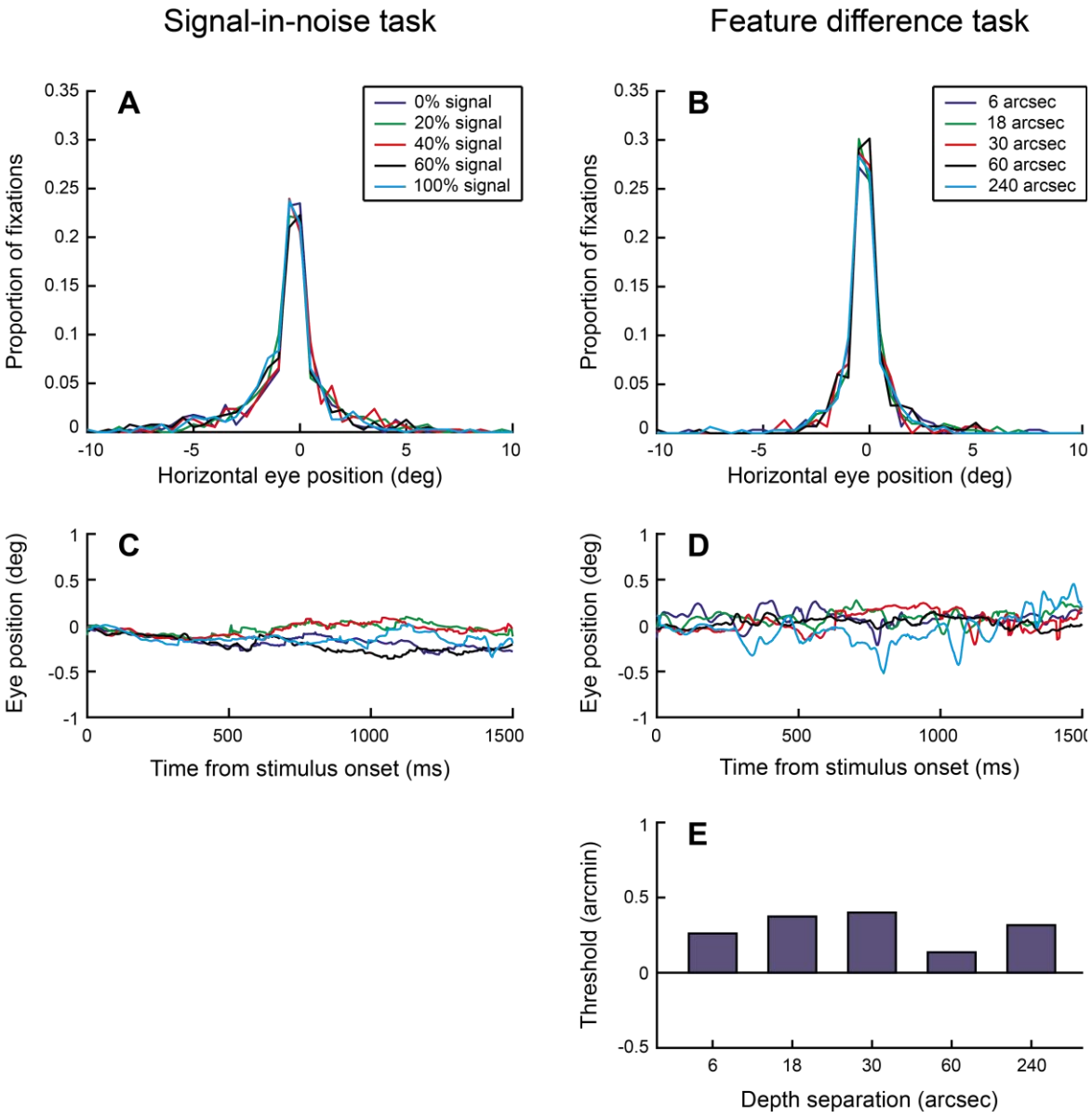


Figure 3.7: Eye-movement analysis and vernier task performance. (A) Horizontal eye movements of three participants that were measured during completion of the signal-in-noise task. This graph shows the proportion of time in each horizontal position where a fixation measurement was recorded, where 0 degrees represents the point of fixation. (B) The corresponding graph for three participants who took part in the feature difference task. (C) Average horizontal eye position from the time of stimulus onset in the signal-in-noise task. (D) The corresponding graph for the feature difference task. (E) Mean threshold values for behavioural performance in the vernier task, displayed simultaneously to stimuli from the feature difference task.

difference between the BOLD signal and the amount of variance observed between trials. However, we found that regions with higher PSC and fSNR were not related to the regions that were observed to have higher prediction accuracies, showing that our results cannot be

explained through changes in the overall fMRI response. Third, to ensure no regions were excluded from our analysis, we conducted a "searchlight" classification analysis (Kriegeskorte, Goebel, & Bandettini, 2006) on all participants for both experiments. This procedure moved a small aperture of 9 mm radius sequentially through the cortex, performing a near-far classification analysis for the stimulus condition with highest signal. This generated a map of prediction accuracies for the whole brain, where voxels with significant prediction accuracies ($t_{(10)} > 2.23$, $p < 0.05$) are displayed in **Figure 3.8**. This confirmed that for both experiments, higher classification accuracies were located in dorsal and parietal visual areas but were also evident in ventral region LO. This confirmed that our definition and selection of ROIs was appropriate. Fourth, it was possible that by training the classifier on the condition with clearest signal, changes in decoding performance as signal level changed (i.e., fMR-metric functions) related to differences in the similarity of the stimuli used for training and testing. Given that we observed differences in the fMR-metric functions between areas, we considered this possibility somewhat unlikely.

Finally, we considered the possibility that changes in activity in the parietal cortex related to task difficulty rather than the processing of disparity signals *per se* (Chandrasekaran et al., 2007). However, we believe such an explanation is unlikely. First, the MVPA technique we used relies on discriminative differences between conditions (i.e., crossed *vs.* uncrossed disparity at the same signal level). We would expect that increases in task difficulty would be expected to have either no influence or enhance discriminative differences between conditions (i.e. the functional purpose of attending), rather than the decline in performance that we observed. Second, our use of a delayed cue paradigm that decoupled the motor response from the perceptual interpretation required that participants maintain task engagement during the delay period, minimising the potential for differences in

attentional allocation for the different signal levels. Therefore, our findings appear more compatible with the perceptual interpretation of disparity information rather than general attentional processes.

3.9 Discussion

We investigated the processing of binocular signals in the human visual system in the context of performance of two different behavioural tasks (signal-in-noise and feature difference tasks). We sought to determine whether these tasks would differentially engage circuits in the dorsal and ventral visual pathways in line with previous suggestions about a division of labour between depth representations in these pathways. We found we were able to decode clearly-defined depth differences in both ventral and dorsal cortical areas; however, the relationship between the decoding of these signals and changes in perception was strongest in higher dorsal stream areas. In particular, our results suggest a close relationship between fMRI responses in parietal regions VIPS and POIPS and the perceived depth for both signal-in-noise and feature difference tasks.

3.9.1 The representation of depth in parietal cortex

The strong association we observed between changes in fMRI activity and perceptual judgments (i.e., fMR-metric *vs.* psychometric functions) suggest that regions of the intraparietal sulcus (VIPS, POIPS) represent depth information at a high level within the chain of computations that starts with matching features between the two eyes and ends with a registered depth impression. Previous work indicated that regions VIPS and POIPS are

activated by a variety of depth configurations, including planar surfaces (Preston et al., 2008), 3D curvature (Georgieva et al., 2009; Minini et al., 2010), depth positions and depth structure (Durand et al., 2009). This is compatible with generalised representations of disparity-defined depth and/or readout mechanisms. Comparisons between monkey and human fMRI suggest that the caudal intraparietal area (CIP) in the monkey could be the homologue of VIPS (Durand et al., 2009; Tsao et al., 2003), or both VIPS and POIPS (Orban et al., 2006a). Single unit recordings from monkey CIP have shown responses to surface structures defined by both binocular disparity (Sakata et al., 1998; Shikata et al., 1996; Taira et al., 2000; Tsutsui et al., 2001; Tsutsui, Sakata, Naganuma, & Taira, 2002) and monocular depth cues (e.g., perspective, texture, motion) (Orban et al., 2006a; Tsutsui et al., 2001; Tsutsui et al., 2002; Vanduffel et al., 2002). Moreover, single unit responses in lateral (LIP) and anterior (AIP) intraparietal areas have revealed responses to depth structure and 3D shape (Durand et al., 2007; Srivastava et al., 2009; Theys et al., 2012b). Further, responses to disparity defined objects were absent for anticorrelated stimuli (Theys et al., 2012b) suggesting high-levels of disparity processing that have rejected false-matches based on a contrast-similarity constraint. These generalised responses are therefore compatible with our evidence that fMRI activity in parietal regions VIPS and POIPS reflects a higher stage of disparity processing that is closely associated with an individual's 3D perceptual interpretation.

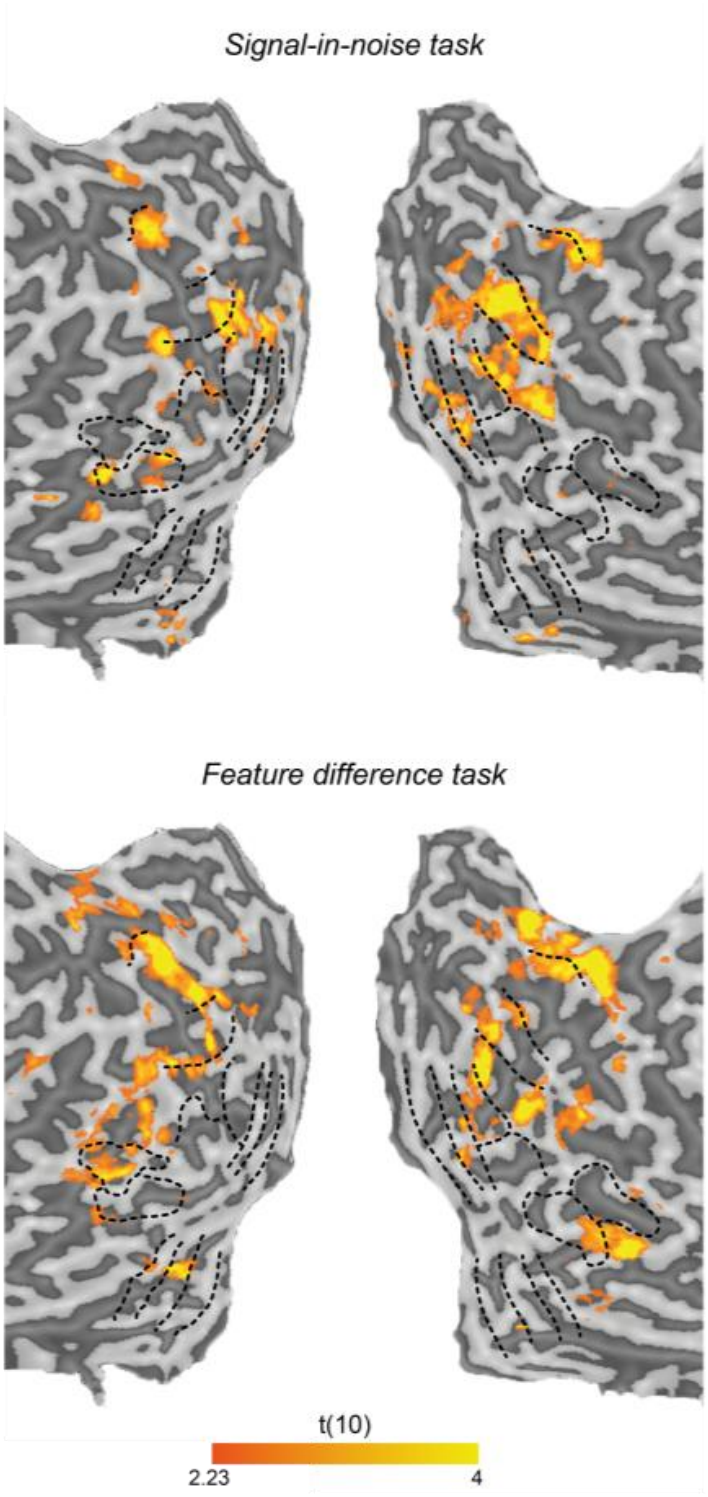


Figure 3.8: Between-subjects results of a searchlight classification. The top row is the average searchlight results from the signal-in-noise task presented on the flatmap of a single participant. The bottom row is the corresponding searchlight results from the feature difference task. All highlighted voxels are significant when contrasted to chance performance (t-test, $p < 0.05$). Both flatmaps are of the same participant, who took part in both experiments. Boundaries of visual areas are identified through dotted lines.

Further, it is possible that the responses we observe in VIPS and POIPS reflect a read-out signal, where computed depth signals are interpreted for functional use. In the context of a motion signal-in-noise tasks, LIP neurons have been shown to be predictive of an animal's perceptual decision (Shadlen & Newsome, 1996). Further, evidence from an ambiguous apparent motion task suggests that neuronal responses in monkey LIP vary in line with the animal's perception (Williams et al., 2003).

3.9.2 A caution about the interpretation of our results

While we observed a close association between perceptual judgments and fMRI responses in VIPS and POIPS, we were also able to decode depth positions in other dorsal areas (V3d, V3A, V3B/KO) as well as ventral area LO. However, we did not obtain strong evidence for an association between psychometric and fMR-metric decoding based performance in these areas. Interpreting these results requires caution for three main reasons. First, a result below the shuffled distribution cannot be taken to imply that a region is not involved in the processing of disparity signals; in particular, our data acquisition and/or analysis methods may be insufficiently sensitive to detect contributions from these regions. Second, an essential component of our paradigm was to measure behavioural performance concurrent to fMRI responses; therefore conditions were required to vary from trial-to-trial. However, results for event-related designs are known to produce noisy estimates due to the successive presentation of different conditions before the hemodynamic response has returned to baseline (Heeger & Ress, 2002). We took measures to reduce the impact of this noise, including counterbalancing conditions and combining the activation from multiple trials prior to classification. Nevertheless, this design and its inherent noise reduced the sensitivity with which we could decode depth for both tasks. In particular, the highest prediction accuracies

we obtained were in the region of 65% which is considerably below the accuracies of up to 85% we have observed elsewhere using similar stimuli in a blocked fMRI design (Ban, Preston, Meeson, & Welchman, 2012; Preston et al., 2008).

Finally, a significant goodness of fit statistic of the psychophysical function to the fMRI data is difficult to interpret when prediction accuracies are near chance. Although prediction accuracies in the lower signal conditions are expected to be located at, or close to chance by nature of the paradigm, it is possible random fluctuations can result in a significant fit of the fMR-metric function (design limitations meant that we were only able to sample 5 points on the function). We minimised the potential for random responses using a criterion that performance in the highest signal case should exceed 99.5% of the accuracies obtained with randomly permuted data in our analysis. However, some regions had prediction accuracies close to chance for all lower signal levels and still contained significant or marginal fits to the scaled behavioural results (for instance, area V7 for the signal-in-noise task; **Figure 3.5**). This was more prevalent for the signal-in-noise task and may be due to the introduction of noise dots distributed across the depth field for both crossed and uncrossed target disparity conditions. The increased similarity between crossed and uncrossed disparities would reduce voxel biases, degrading classification performance.

3.9.3 Neural responses in earlier dorsal and ventral visual areas

Responses in early visual areas (V1, V2) did not exceed the shuffling distribution for either task (**Figure 3.4**) and this is consistent with the notion that these areas relate to the processing of local disparity signals rather than the global percept (Bredfeldt & Cumming, 2006; Chandrasekaran et al., 2007; Cumming & Parker, 1997, 2000; Cumming, Shapiro, & Parker, 1998; Preston et al., 2008).

We provide suggestive evidence for the role of earlier dorsal regions in both tasks. We revealed that earlier dorsal regions (V3A, V3B/KO) may be associated with psychophysical performance for both the signal-in-noise (**Figure 3.5**) and feature difference tasks (**Figure 3.6**). However, prediction accuracies for these areas showed only limited sensitivity and we are therefore cautious in suggesting an involvement in these regions for our tasks. Our statistics also suggested a good fit of the behavioural results in V7 for the signal-in-noise task; however visual inspection of the data revealed a result analogous to LO, where the introduction of disparity noise reduced classification performance to chance. Interestingly, we did not observe significant decoding accuracies in hMT+/V5 for either task, even though single-unit recording has been performed extensively in this region during coarse disparity tasks (Chowdhury & DeAngelis, 2008; DeAngelis et al., 1998; Uka & DeAngelis, 2003, 2004, 2006). A salient factor is that we presented static stimuli, whereas previous work has used moving stimuli containing disparity differences. It is therefore possible that we might have seen more involvement of hMT+/V5 if our stimuli had involved conjunctions of motion and disparity.

In contrast, it was only LO in the ventral visual pathway that showed responses that could potentially be associated with depth judgments. The ventral pathway is thought to encode fine disparity differences, and our results for the feature difference task provided suggestive evidence towards this role. This supports previous work that has observed responses to depth stimuli in LO at the population level for both depth planes (Preston et al., 2008) and 3D shape (Chandrasekaran et al., 2007). Similarly, neurons in a later part of the ventral stream, the inferior temporal cortex, were seen to directly reflect perception of 3D shape in the monkey (Janssen et al., 2003; Janssen et al., 2000a). For the signal-in-noise task, responses of LO reduced to chance with the addition of any noise, even when participants

were capable of discriminating depth position. One possibility for this result is that ventral regions are unable to detect signal-from-noise, since high performance on the 100% signal condition is without noise and simply reflects a feature difference task.

3.9.4 The effect of dot density on neural processing

In ideal circumstances, humans are perceptually able to distinguish very small changes in depth (~5 arcsec) (McKee, 1983; McKee et al., 1990; Stevenson et al., 1989). However, a psychometric function generated for values about this threshold would cover only a small range of disparities and the MVPA classifier would be unlikely to show parametric variation between stimulus conditions. Thus, it was necessary to increase the difficulty of the task through other stimulus properties so that we were able to use a larger range of disparity magnitudes. This was accomplished by using a pedestal that was located away from the plane of fixation and by lowering the dot density.

However, it is possible that the change in dot density could still have affected our results. Although changes to dot density of random-dot stereograms can be used to affect psychophysical judgments (Cogan, Lomakin, & Rossi, 1993; Cormack, Landers, & Ramakrishnan, 1997), this does not generally appear to be reflected in the responses of individual neurons. For example, changing dot density from 25% to 50% in disparity-defined 3D shapes did not affect neuronal responses in the inferotemporal cortex (Janssen, Vogels, Liu, & Orban, 2001; Janssen et al., 2003). Similarly, changing the dot density between 10 – 20% affected the amplitude of responses in some V1 neurons, though it did not change the overall disparity preferences of the neuron (Trotter, Celebrini, Stricanne, Thorpe, & Imbert, 1992, 1996). Further, in terms of random-dot motion patterns, changes were only observed in the response of V1 and MT neurons for very low densities (below ~20 dots) and overall dot

density was suggested to have little effect on neural responses (Snowden, Treue, & Andersen, 1992; Snowden, Treue, Erickson, & Andersen, 1991). Thus, we believe it is unlikely that the change in dot density between our experiments had an impact on our observed results.

3.9.5 The relation to results from coarse and fine disparity tasks

Many studies that have found a distinction between coarse and fine disparities between the dorsal and ventral visual pathways have done so by using two separate tasks: coarse disparities are investigated by discrimination of depth position from a random-dot stereogram that is comprised of varying proportions of noise, while fine disparities are investigated through discrimination of small differences between two disparity planes. However, both the magnitude of the disparity and the perceptual task has changed. By examining cortical responses to these two tasks while limiting the range of disparity magnitudes presented, we have decoupled these two features. We found dorsal visual areas, particularly VIPS and POIPS, mirrored perceptual judgments for both signal-in-noise and feature difference tasks. These results indicate that it is not the perceptual task that is used which creates a distinction between the visual pathways, and that both tasks require processing in the dorsal visual pathway.

However, the response we observed in parietal regions for both perceptual tasks contrasts with a previous study that found microstimulation in MT was not observed to bias performance on a fine disparity task (Uka & DeAngelis, 2006). However, there are several possible reasons for this difference. First, it could represent a difference in disparity processing between the monkey and human visual system. Second, neuronal responses in MT are unlikely to be representative of all dorsal visual areas in humans. Our findings in parietal regions VIPS and POIPS may reflect a later stage of processing that does not rely on the

response of MT neurons, as monkey LIP has been observed to receive inputs from V3 (Adams & Zeki, 2001) and V3A (Nakamura et al., 2001). This activation is then projected to anterior parietal regions for use in visually-guided actions (Nakamura et al., 2001), consistent with our suggested role of VIPS and POIPS. Third, our use of fMRI measures the population response of each voxel, rather than preferences of individual neurons. Therefore, neurons selective to disparity in these regions may be distributed more sparsely and are concealed by the response of many other neurons that were not sensitive to our task.

3.9.6 Conclusion

In summary, we simultaneously measured fMRI activity and behavioural performance on a depth discrimination task and used a multivariate classifier to identify cortical regions involved in the perception of disparity-defined depth. This was performed on signal-in-noise and feature difference tasks, commonly used in the study of coarse and fine disparities, to characterise responses in both ventral and dorsal pathways in the human visual cortex. Our study revealed higher sensitivity in the dorsal stream and in particular, parietal regions VIPS and POIPS in the processing of perceptually-relevant signals for both of our tasks. The likely role of these regions is to integrate depth information and form a generalised representation of 3D surfaces. Future studies could use TMS to identify whether the visual regions identified here play a causal role in judgments of disparity-defined depth.

4. Stereo correspondence and binocular perception

Abstract

Stereo correspondence requires matching the features from one eye to the corresponding features in the other eye. Only after this matching process takes place can disparity be computed and binocular depth perceived. We investigated cortical regions in the human visual system that had solved the correspondence problem by manipulating behavioural performance and identifying visual areas that varied in line with depth judgments. We manipulated task difficulty by presenting random-dot stereograms with varying proportions of anticorrelated dots and recorded event-related fMRI responses obtained concurrently. We compared changes in behavioural performance to the corresponding changes in performance of a multivariate classifier. We found that in the dorsal stream, regions V7 and VIPS had solved the stereo correspondence problem and were highly diagnostic of perceptually useful, disparity-defined depth.

4.1 Introduction

The horizontal separation of our two eyes causes slight differences in the view of each eye. These differences, known as binocular disparities, are used to perceive depth in our environment. To achieve stereopsis, a feature in the image of one eye must first be matched with the corresponding feature in the other eye (i.e., the correspondence problem; Marr & Poggio, 1979). Only then can the binocular disparity be computed to reveal the depth of the object. This matching process is a computationally difficult task that the human visual system performs seemingly effortlessly (Harris, 2004), however relatively little is known about its underlying neural circuitry and how this relates to visual perception.

To understand the difficulty of the correspondence problem, consider that for correlated random-dot stereograms (RDS) (**Figure 4.1a**, right image) a given dot in one eye has a host of possible matches in the other eye. Most of these are false matches and there is only one pairing that will lead to a globally consistent solution (Julesz, 1971). Finding this match is made more challenging as the difference in the inter-ocular position of the feature is dependent on the depth of the object. The visual system must therefore discard incorrect choices while preserving the correct match that is in agreement with the global percept. Spatial filtering of the image reduces the computational burden (Marr & Poggio, 1979), but is insufficient to find the appropriate match without further processing.

Responses to binocular disparity have been observed throughout the visual cortex (for a review; DeAngelis, 2000; Neri, 2005; Parker, 2007), but this does not identify their role in binocular tasks. Further, it does not separate whether these responses have overcome the correspondence problem or are simply the result of local matches at an early stage of processing. Features have to be paired in several characteristics for a match to take place, including comparison of (i) luminance values, (ii) edge discontinuities, and (iii) epipolar

position such that each point has a unique location in space. Specific manipulation of luminance is performed with anticorrelated RDS (**Figure 4.1a**, left image), where the contrast of dots in one of the eyes is reversed (i.e., a black dot in one eye is paired with a white dot in the other eye and *vice versa*) and depth is no longer perceived (Cogan et al., 1993; Cumming et al., 1998; Read & Eagle, 2000). The binocular energy model predicts that monocular subunits with opposite polarity respond with an inverted tuning curve (Cumming & Parker, 1997; Ohzawa et al., 1990; Read et al., 2002) and this is indeed observed for cells in the primary visual cortex (Cumming & Parker, 1997). However, as anticorrelated RDS do not support depth judgments, these false matches must be filtered out and rejected by the visual system.

Such responses are suppressed by the endpoints of both the dorsal and ventral visual pathway (Janssen et al., 2003; Theys et al., 2012b), however it remains unclear where the neural response begins to reflect perceptual judgments in either stream. For the ventral stream, responses to anticorrelated RDS are diminished in intermediate region V4 (Kumano et al., 2008; Tanabe et al., 2004) while for the dorsal stream many neurons are still responsive in later dorsal regions MT and MST (Krug et al., 2004; Takemura et al., 2001). By comparison, human neuroimaging has revealed a preference towards correlated over anticorrelated RDS in all measured dorsal regions, while this was only observed in ventral region LO (Preston et al., 2008).

In this study, rather than compare performance between correlated and anticorrelated RDS, we manipulated behavioural performance by varying the proportion of anticorrelated dots (e.g., **Figure 4.1a**) on a depth discrimination task. We then sought to identify cortical regions which responded in line with the observers' depth percept. As in Chapter 3, behavioural performance was measured concurrent to fMRI measurements and we used a

multivariate classifier to determine the extent to which depth position could be decoded in pre-defined regions of interest. We compared changes in behavioural performance to the corresponding changes in classification performance to suggest areas that related to binocular depth perception.

4.2 Methods

4.2.1 Participants

Eleven participants (five females) naïve to the purposes of the experiment were recruited from the University of Birmingham. Mean age was 22.6 years (range, 18 – 33 years). All participants had normal or corrected-to-normal vision with no deficits in colour vision and were paid for their participation. Prior to scanning, all participants were screened for stereo deficits in the laboratory haploscope to ensure stimuli presented in the study were above observers' detection thresholds. One additional observer was excluded from the data analysis due to excessive head movement during scanning (greater than 7 mm or 7 degrees drift over the course of the scan). The experiment was granted ethics approval and all participants provided written informed consent.

4.2.2 Stimuli

We used random dot stereograms (RDS) defined by black and white dots within a rectangular aperture (14 × 19°) and displayed on a mid-grey background. Within this region, a central test plane (10 × 10°) was given a nonzero disparity of ±6 arcmin relative to the fixation plane. To minimise the effects of adaptation, some jitter (up to ±1 arcmin) was added to this disparity value for each trial. The dot density of the stereogram was 2 dots/deg² and each dot had its

edges softened using a Gaussian distribution (diameter 0.15° at half-height). The background surrounding the RDS was filled with a grid of black and white squares which could be used as a clear reference and encouraged vergence to the plane of the screen. The fixation marker was a hollow white square (0.5° side length) that was presented in the centre of the stereogram. Vertical and horizontal nonius lines (length 0.375°) were attached to the fixation marker to promote vergence at the plane of the screen. We restricted presentation of dots to outside of a circular region (1.5° diameter) centred at fixation to ensure the stimulus did not interfere with binocular fusion.

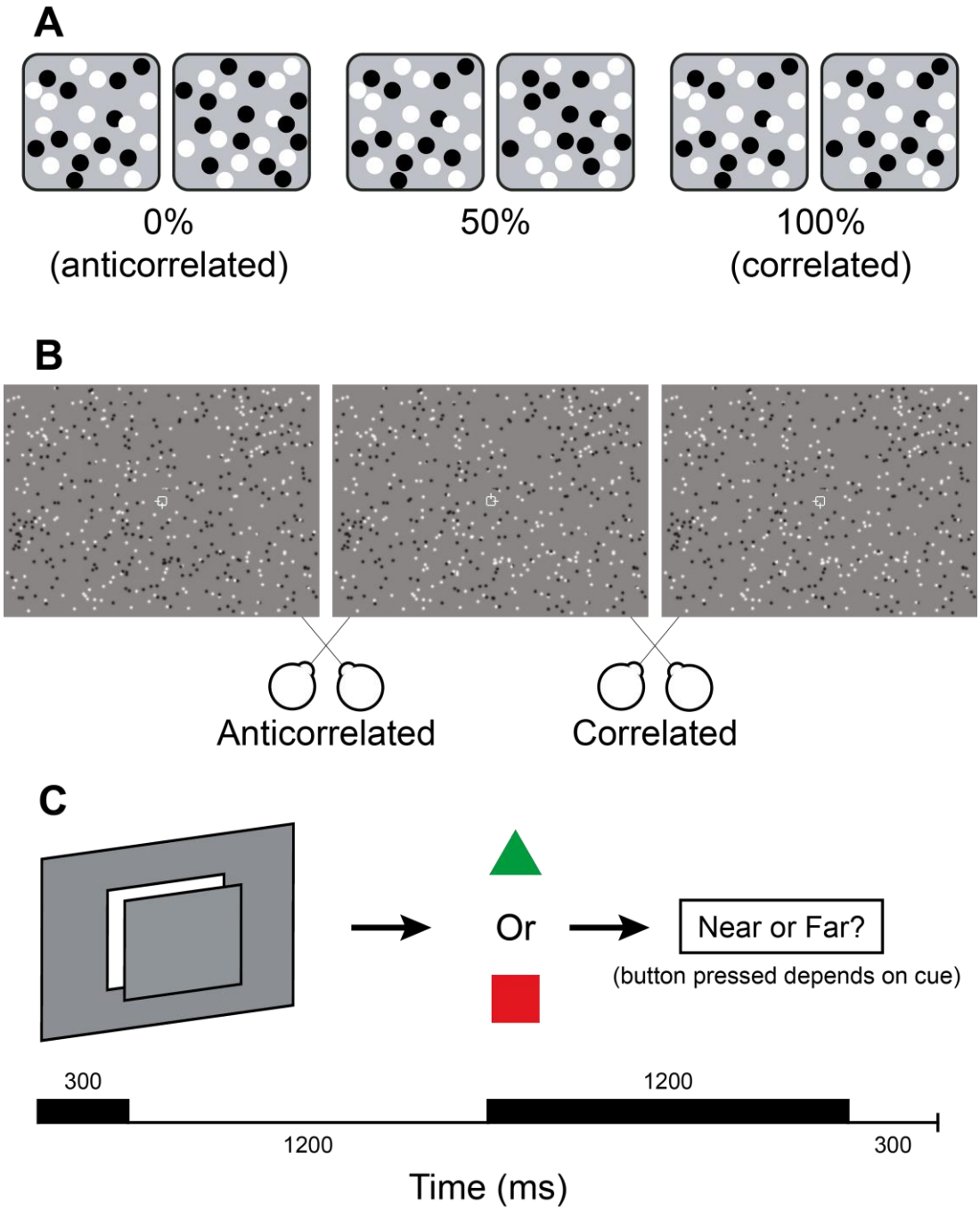


Figure 4.1: (A) A cartoon RDS that shows three proportions of correlated dots. Each box represents the image presented to one of the eyes. In the 100% correlated condition (right), the dots are perfectly correlated and every dot in one eye is paired with a dot of the same contrast in the other eye. In the 0% correlated condition (left), the dots are perfectly anticorrelated and every dot in one eye is paired with a dot of the opposite contrast in the other eye. The 50% correlated condition (middle) contains an equal proportion of correlated and anticorrelated dots. (B) An example of correlated and anticorrelated stimuli used in the experiment. The middle and right images are perfectly correlated and can be free-fused to show a square with an uncrossed disparity. The left and middle images are anticorrelated and do not contain a depth percept. (C) The procedure for a typical trial. The first black bar indicates stimulus onset and duration, and the second black bar indicates presentation of the response cue.

We manipulated task difficulty by varying the contrast correlation of dots presented to the two eyes for the target RDS. At 100% correlation, white dots in one eye matched white dots in the other, and black dots corresponded to black dots. However, at 0% (perfectly anticorrelated) white dots in one eye were matched to black dots in the other. We generated five conditions that expressed the proportion of dots in the display that were correlated: 0, 20, 40, 60 and 100% (i.e. a lower number implies a higher proportion of anticorrelated dots; **Figure 4.1a**). All dots in the RDS surrounding the central test plane were located at the plane of fixation and were perfectly correlated. An example of the stimulus in both correlated and anticorrelated conditions is presented in **Figure 4.1b**.

4.2.3 Design

For each participant, we collected both behavioural and fMRI data concurrently in a single session of approximately nine event-related scans. In each trial, participants made a perceptual judgment on the depth sign of the stimulus (i.e., 'near' or 'far'), and were trained in this task prior to scanning. Trials were randomised and counterbalanced such that each condition preceded all other conditions equally (for one trial back), with the order of trials differing between runs and participants. Eleven conditions (ten stimulus conditions and one fixation condition during which the central test plane was removed) with 11 trials per condition were presented in each run. Each run comprised 122 trials (121 across conditions and one initial trial for balancing the history of the second trial) and two 9 s fixation periods (one in the beginning and one at the end of the run), lasting 6 min 24 s. The ten conditions consisted of the five noise levels for both crossed and uncrossed disparities.

For fixation trials, the fixation square and surround RDS was displayed for 3 s. Experimental trials lased 3 s (**Figure 4.1c**) and started with a stimulus presentation of 300 ms

followed by a delay of 1200 ms (75% of trials) or 1400 ms (25% of trials) during which the central test plane was removed from the screen. These different delay times were chosen to minimise predictability and habitual responding by participants. After the delay, a green triangle or red square appeared inside the fixation marker and served as an indicator for the motor response mapping to be used on that trial. In particular, if the response cue was a green triangle observers used a particular finger-key matching (e.g., index finger for 'near'), while if the response cue was a red square, observers reversed the finger-key matching (e.g., index finger for 'far'). This was balanced across participants to remove any bias for a particular cue. The response cue was removed 300 ms before the start of the next trial. This procedure separated the motor response (i.e., button press) from the perceptual depth judgment. Participants were familiarised with this process in the laboratory prior to scanning.

4.2.4 fMRI data acquisition

The study was performed in a 3-Tesla Philips Achieva MRI scanner at the Birmingham University Imaging Centre. We used an eight-channel SENSE head coil to collect echo-planar imaging (EPI) and T1-weighted anatomical (1 × 1 × 1 mm) data. For experimental runs, EPI data [echo time (TE), 35 ms; repetition time (TR), 1500 ms] were obtained from 25 slices (voxel size, 2 × 2 × 2 mm, 256 volumes) positioned close to the coronal plane. Localisers were obtained in a separate session, with EPI data (TE, 34 ms; TR, 2000 ms) acquired from 28 slices (voxel size, 1.5 × 1.5 × 2 mm, near coronal). We used a pair of video projectors (JVC D-ILA SX21) to display stereoscopic images, with each projector fitted with a unique interference filter (INFITEC, GmBH) that distributed the wavelengths of visible light between the two projectors. The images from each projector were optically combined using a beam-splitter cube and passed through a waveguide into the scanner room. Stimuli were projected

onto a translucent plastic screen located behind the head coil and inside the bore of the scanner. Participants viewed the screen via a mirror positioned on the head coil angled at 45°, with an optical path length of 65 cm. Since the INFITEC filters produced negligible overlap between the spectra for each projector, there was minimal crosstalk between the projected images for a participant wearing a pair of corresponding filters. Unique stimuli were pre-generated for each participant using C#, and the experiment was performed using MATLAB (The MathWorks, Natick, MA) and the PsychToolBox 3 extension (Brainard, 1997; Pelli, 1997).

For each participant, we identified regions of interest (ROIs) from independent data in a separate localiser scan. We used a rotating wedge stimulus and expanding concentric rings to define the borders of early retinotopic areas (V1, V2, V3v, V4, V3d, V3A and V7) (Aguirre et al., 1998; DeYoe et al., 1996; Sereno et al., 1995). In particular, V4 was defined as the ventral region adjacent and inferior to V3v containing a full hemifield representation (Tootell & Hadjikhani, 2001; Tyler et al., 2005) and V7 was defined as the dorsal region adjacent and anterior to V3A, also containing a full hemifield representation (Tootell et al., 1998; Tsao et al., 2003; Tyler et al., 2005). We identified higher dorsal regions [V3B/kinetic occipital area (KO), human motion complex (hMT+/V5)] and the ventral lateral occipital region (LO) from additional localiser scans. Area V3B/KO (Dupont et al., 1997; Zeki et al., 2003) was defined anatomically as the dorsal region adjacent and lateral to V3A with which it shared a foveal representation (Tyler et al., 2005), and functionally as the set of voxels that responded significantly more ($p < .001$) to kinetic boundaries than to the transparent motion of black and white dots without clear borders. Area hMT+/V5 was defined as the region in the lateral temporal cortex that responded significantly more ($p < .001$) to an array of coherently moving dots than to an array of static dots (Zeki et al., 1991). The lateral occipital (LO) area was

identified as the region in the lateral occipito-temporal cortex that responded significantly more ($p < .001$) to images of objects and shapes than to scrambled versions of these images (Kourtzi & Kanwisher, 2000, 2001). During all localiser scans, participants performed an attentionally demanding task on the fixation point, except for the LO localiser in which they had to respond if the same image was presented consecutively. Finally, the ventral (VIPS) and parieto-occipital (POIPS) regions along the intraparietal sulcus were defined as regions anterior to V7 which showed significantly stronger ($p < .001$) responses to all of the experimental conditions in contrast to the fixation baseline. For illustrative purposes, the mapping of all these ROIs for one participant is presented in **Figure 4.2**.

4.2.5 fMRI data analysis

We processed MRI data using BrainVoyager QX (BrainInnovation, Maastricht, the Netherlands). For each participant, anatomical scans were transformed into Talairach space and used for 3D cortex reconstruction, inflation, flattening and the segmentation of gray and white matter. Preprocessing of functional data included head movement correction (translation and rotation), slice scan time correction, and removal of low-frequency drifts and linear trends through temporal high-pass filtering (3 cycles). Functional runs were aligned to the participant's anatomical data and then transformed into Talairach space. Consecutive volumes that contained head movement greater than 1 mm of translation or 1° of rotation were excluded from further analysis.

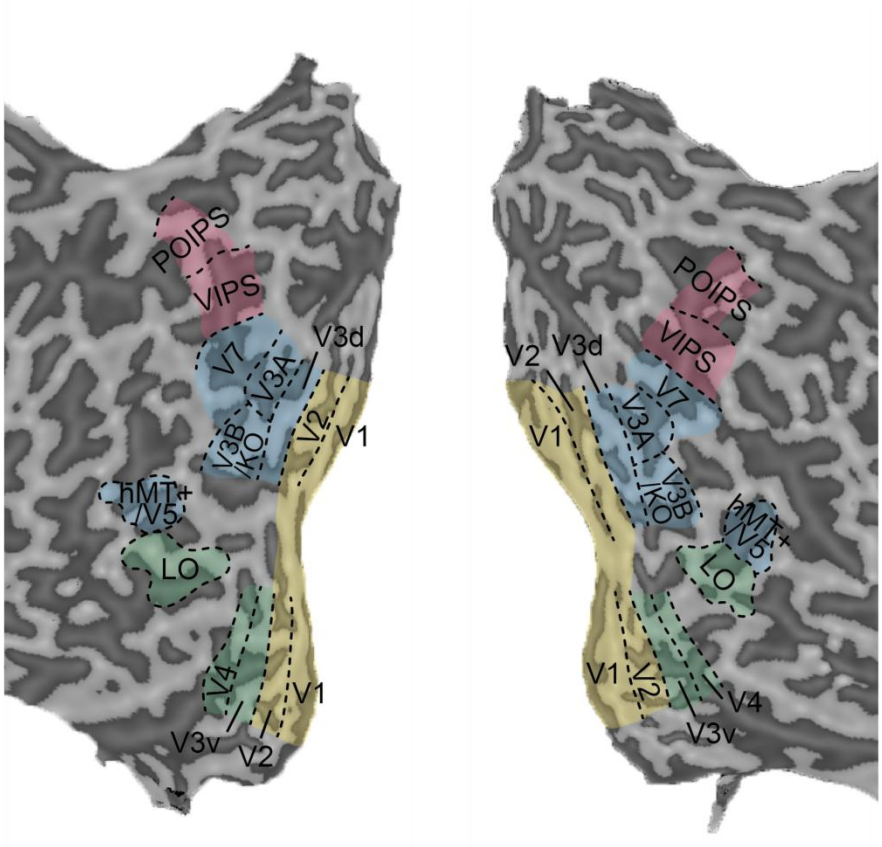


Figure 4.2: Regions of interest in one participant showing retinotopic areas, V3B/KO, hMT+/V5, LO and parietal regions VIPS and POIPS.

We used Multi-Voxel Pattern Analysis (MVPA) software to analyse the fMRI data from each ROI. For voxel selection, gray matter voxels were isolated through a grey-matter mask from the anatomical scan and only included voxels that showed a *t* value larger than 0 when contrasting fMRI activity from all stimulus conditions to the fixation baseline across all experimental runs. These were then sorted according to their response (*t* statistic) and the 300 most active voxels (or highest number available) in each ROI were used in the classification. We excluded V3A from the analysis in one participant as it contained less than 75 voxels after this contrast was applied. We were also unable to obtain data for the parietal region POIPS for two participants due to the spatial resolution of the EPI sequence and near coronal slice positioning during the fMRI acquisition process. Estimation of fMRI responses to single

events in our event-related fMRI design were likely to be noisy for single trials; therefore, prior to feeding the data to the machine learning classifier we averaged a small number of trials (4, 4 and 3 from a single run) to generate 3 training patterns per run. Each voxel's time series was then normalised (z-score) in each experimental run to compensate for the decrease in mean image intensity with distance from the receiver coil. The fMRI time series was shifted by 3 volumes (4.5 s) to account for the hemodynamic delay of the BOLD signal. Each volume had the mean univariate signal subtracted to remove any bias created from it.

We used a leave-one-run-out cross-validation sampling procedure, to ensure generalisation of the classification. For each cross-validation, one run was left out as an independent test dataset and the data from the rest of the runs was used as the training set. The classification accuracy for each ROI was obtained by averaging the prediction accuracy across cross-validations. Training was performed on the 100% signal condition, and test patterns were calculated for all five signal levels. The reported prediction accuracy of the classifier corresponds to the proportion of trials on which it could correctly predict the stimulus based on the pattern of fMRI responses, in which chance performance would be 0.5 for a binary classification (i.e., 'near' *vs.* 'far' stimuli). For each ROI, prediction accuracies were averaged across all participants. Plotting prediction accuracy across pattern size (number of voxels) showed that classification values had saturated by 300 voxels, validating our choice of pattern size.

4.3 Results

Participants were asked to judge the depth position of a test plane while we simultaneously measured fMRI activity in the visual cortex. We manipulated the difficulty of the task by increasing the proportion of dots in the random-dot stereogram that were anticorrelated and therefore did not support a perceptual interpretation of stereopsis. This reduced the proportion of correlated (i.e., signal) dots that could be used to determine the target disparity. We chose five levels of the percentage of correlated dots to sample the psychometric function and investigated whether changes in the perceptual interpretation of the stimuli were mirrored by changes in the performance of the classifier. We fit a cumulative Gaussian to the mean

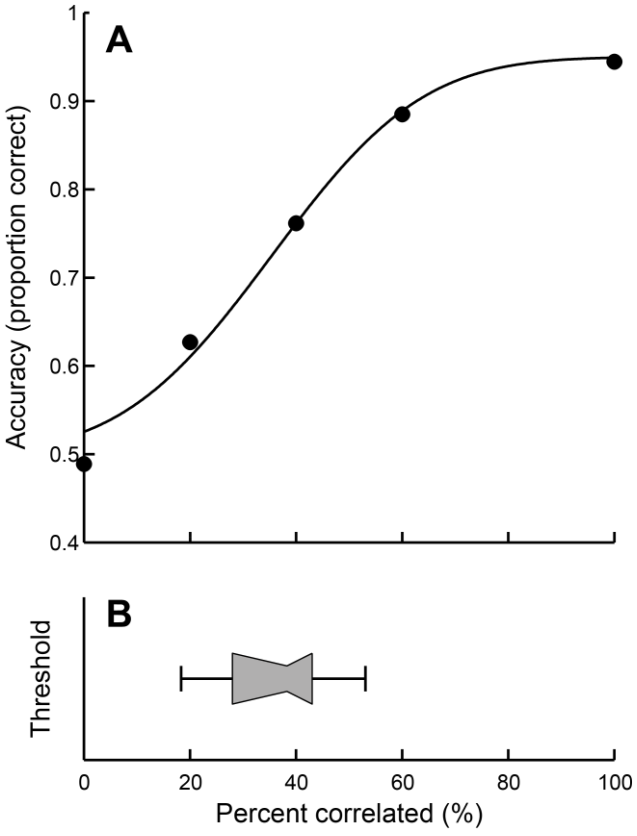


Figure 4.3: Behavioural results of the depth discrimination task. (A) Mean task performance where accuracy refers to the proportion of responses that participants correctly identified as near or far. The solid curve is the best-fitting Gaussian. (B) Distribution of threshold values for the discrimination task. The centre of the 'bowtie' represents the median, the shaded area depicts 68% confidence values, and the error bars are the 95% confidence intervals.

behavioural performance, finding that observers required, on average 35.41% of dots to be correlated to reliable discriminate near from far depth positions (**Figure 4.3**). This level of performance is consistent with thresholds observed from other laboratories (Doi, Tanabe, & Fujita, 2011).

4.3.1 fMR-metric functions to anticorrelated RDS

We investigated how classification performance varied as a function of the proportion of correlated dots for visual areas that have previously been implicated in disparity processing (Backus et al., 2001; Minini et al., 2010; Neri et al., 2004; Rutschmann & Greenlee, 2004). We trained a linear support vector machine to associate patterns of voxel activity within each visual area to the disparity-defined depth position of the stimulus that gave rise to the activity. We tested whether we could predict the viewed stimulus from the fMRI activity, calculating the mean leave-one-run-out prediction accuracy for classifiers trained to discriminate crossed from uncrossed disparities when no noise was present. We compared the results of the classifier to the mean behavioural performance obtained simultaneously to identify cortical

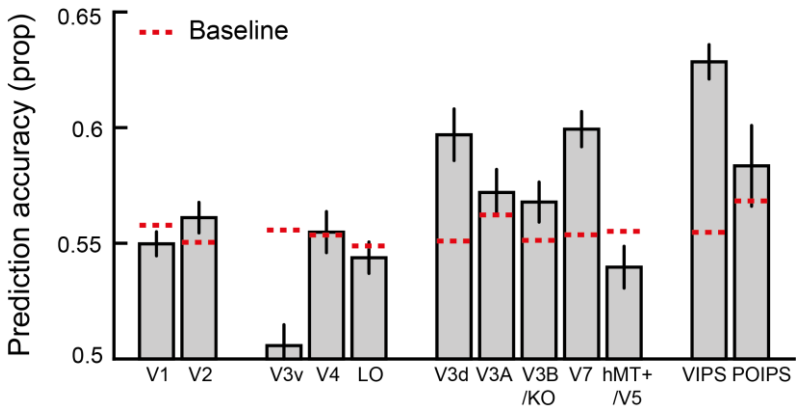


Figure 4.4: Mean prediction accuracy of the classifier for the 100% correlated condition. Horizontal red lines mark the baseline of statistical significance generated from randomly permuting data labels before being fed into the classifier. Error bars are standard error.

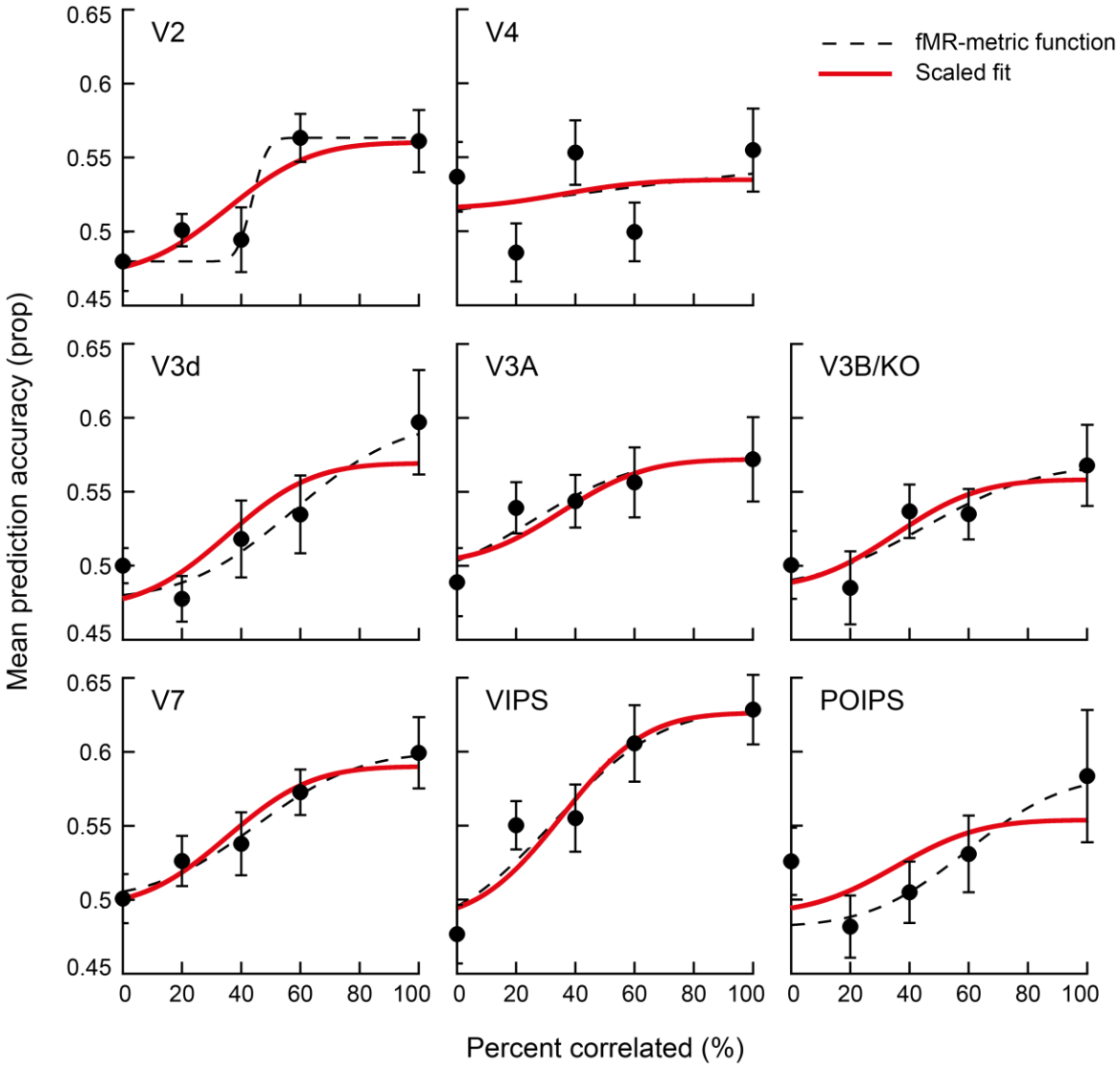


Figure 4.5: Mean fMR-metric functions for each visual area. The red line is a scaled version of the behavioural results and the dotted line is the best-fitting Gaussian. We have presented only visual areas that performed significantly above chance. Error bars are standard error.

regions that responded in a manner that mirrored changes in the perceptual discriminability of the stimuli.

Figure 4.4 shows the mean prediction accuracies for the 100% correlated condition for all ROIs. To determine the level of classification performance we would expect by chance, we ran the classification analysis with randomly permuted fMRI patterns (i.e., we randomised the correspondence between fMRI data and training labels and estimated the classifier

prediction for each visual area) over 999 iterations for the 100% correlated condition. This created a distribution of classification accuracies which we used as a criterion for statistical significance (Bonferroni corrected; **Figure 4.4**, dotted lines). For all regions of interest, the median of the shuffled distribution was at chance (range, 0.499 – 0.501); this confirmed that our analysis technique was unbiased. With the exception of hMT+/V5, all measured dorsal and parietal visual areas supported classification accuracies that exceeded the criterion for random behaviour (the mean prediction accuracy of the 100% correlated condition was above the upper 99.5% centile of the shuffled distribution). Of the other areas considered, only early visual area V2 and ventral region V4 marginally passed this threshold.

To test whether there was a relation between the participants' perceptual judgments and our ability to decode the stimulus from fMRI activity, we calculated the performance of the classifier at different proportions of anticorrelated dots. We fit this with a cumulative Gaussian to generate 'fMR-metric' functions for each ROI. To compare classifier performance to that of the participants, we used the parameters (mean, threshold) of the psychometric function (**Figure 4.3a**) to constrain a Gaussian fit to the fMRI data, allowing only the maximum and minimum values to vary as free parameters. This created a scaled version of the behavioural results and allowed us to make direct comparisons between the simultaneously recorded fMRI activity and behavioural performance. We present these fMR-metric functions for ROIs with above chance performance in **Figure 4.5** and performed a χ^2 goodness-of-fit test to quantify the fit of the fMRI classification accuracies to the values on the scaled fMR-metric function (**Table 4.1**). fMRI responses in the majority of dorsal visual areas (V3A, V3B/KO, V7, VIPS) as well as early visual region V2 could be significantly decoded in a manner that was similar to the behavioural performance. Responses in intermediate dorsal region V3d ($r = .856$, $p = .064$) were statistically marginal, while we did

	Pearson Corr	P-value
V2	.905	**.034***
V4	.246	.690
V3d	.856	.064
V3A	.902	**.036***
V3B/KO	.900	**.037***
V7	.973	**.005***
VIPS	.950	**.013***
POIPS	.665	.220

Table 4.1: Goodness-of-fit of the fMR-metric functions. A χ^2 goodness-of-fit test of the fMRI data points (Fig. 5) to the scaled version of the behavioural results (Fig. 3). An asterisk marks a significant fit for that ROI ($p < .05$).

not observe a similar parametric effect in ventral region V4 ($r = .246$, $p = .690$) or parietal region POIPS ($r = .665$, $p = .220$), suggesting a mismatch between changes in the fMRI response and the perceptual response.

To ensure that we were not missing any important loci of cortical activity related to the task, we conducted a "searchlight" classification analysis (Kriegeskorte et al., 2006) for all participants. This procedure moved a small aperture of 9 mm radius sequentially through the cortex, performing a near-far classification analysis on the 100% correlated RDS. We thus generated a near-far prediction accuracy map for the whole brain (**Figure 4.6**). This revealed that higher classification accuracies were centred on later dorsal visual areas V7 and VIPS, confirming the results we obtained in the fMR-metric functions using our region of interest approach. Interestingly, we also noticed a small area medial to VIPS that was activated by our searchlight analysis. This was located posterior to the anatomical definition of V6 that is involved in optic flow and self motion (Pitzalis et al., 2010). It is possible that the activation is located on the boundary of regions IPS1-4, as defined by a delayed saccades paradigm (Sereno, Pitzalis, & Martinez, 2001; Swisher, Halko, Merabet, McMains, & Somers, 2007). However, one of our participants was localised with these parietal regions and the searchlight results were still observed to be medial of this definition.

Our results highlight a role of later dorsal regions V7 and VIPS in the perception of disparity-defined depth, where we observed a close association between classification performance and the behavioural percept of the observer. In contrast, the prediction accuracies observed for earlier visual regions (V2, V3A, V3B/KO) were close to chance and thus we were unable to clearly indicate a similar association for these areas. Previously, Preston *et al.* (2008) used human fMRI and multivariate techniques to compare the difference between perfectly correlated and anticorrelated stimuli. The authors successfully decoded anticorrelated RDS in several of these regions and observed similar prediction accuracies in V2 and V3d for both stimulus types. We therefore would not expect to find an fMR-metric function in these regions. Instead, responses in these earlier dorsal regions are likely the result of noise. Sensitivity in our paradigm was likely reduced from this previous study due to our use of an event-related design (needed to capture behavioural responses on the depth discrimination task). We found responses to correlated stimuli across visual areas at a maximum of ~65% while previous studies that used similar stimuli and disparity magnitudes in a block-design reached accuracies of ~85% (Ban et al., 2012; Preston et al., 2008). These results support previous studies that have illustrated sensitivity of disparity-defined depth in dorsal regions (Backus et al., 2001; Cottereau et al., 2011, 2012a; Minini et al., 2010; Preston et al., 2008; Tsao et al., 2003), while also revealing a close link between perceptual and cortical responses in these areas.

4.3.2 Control analyses

We performed several control measures and designed our stimulus to ensure that the results were best explained through changes in the perceptual depth interpretation. First, we recorded horizontal eye-movements for three participants using a monocular limbus eye-tracker (CRS

ltd). Analyses of these data suggested no systematic difference in eye-position (**Figure 4.7**) and no statistical difference in the number of saccades between conditions, suggesting it was unlikely that eye-movements could adequately explain our results. The confines of the fMRI scanner and our use of spectral filters for stereoscopic presentation meant that we were unable to record changes in eye vergence; however we designed our stimuli to minimise this possibility. In particular, we instructed participants to maintain alignment of vertical and horizontal nonius lines that surrounded the fixation marker throughout all runs, encouraging vergence at the correct depth (the plane of the screen). Further, the stereogram was surrounded by a clear reference pattern that was located at the fixation plane.

Second, to ensure our results were related to disparity processing and not an overall change in the univariate signal, we computed the percent signal change (PSC) by comparing the difference in the BOLD signal between the 100% correlated condition and fixation in each ROI. Similarly, we examined the functional signal-to-noise ratio (fSNR) by comparing the difference between the BOLD signal and the amount of variance observed between trials. However, we found that regions with higher PSC and fSNR were not related to the regions that were observed to have higher prediction accuracies, showing that our results cannot be explained through changes in the overall fMRI response. Third, it was possible that by training the classifier on the 100% correlated condition, changes in decoding performance as signal level changed (i.e., fMR-metric functions) related to differences in the similarity of the stimuli used for training and testing. However, we observed differences in the fMR-metric functions between visual areas, making this possibility somewhat unlikely.

Finally, we considered the possibility that changes in activity in the parietal cortex related to task difficulty rather than the processing of disparity signals *per se* (Chandrasekaran et al., 2007). However, we believe such an explanation is unlikely. First, the MVPA technique we used relies on discriminative differences between conditions (i.e., crossed *vs.* uncrossed disparity at the same signal level). We would expect that increases in task difficulty would be expected to have either no influence or enhance discriminative differences between conditions (i.e. the functional purpose of attending), rather than the decline in performance that we observed. Second, our use of a delayed cue paradigm that decoupled the motor response from

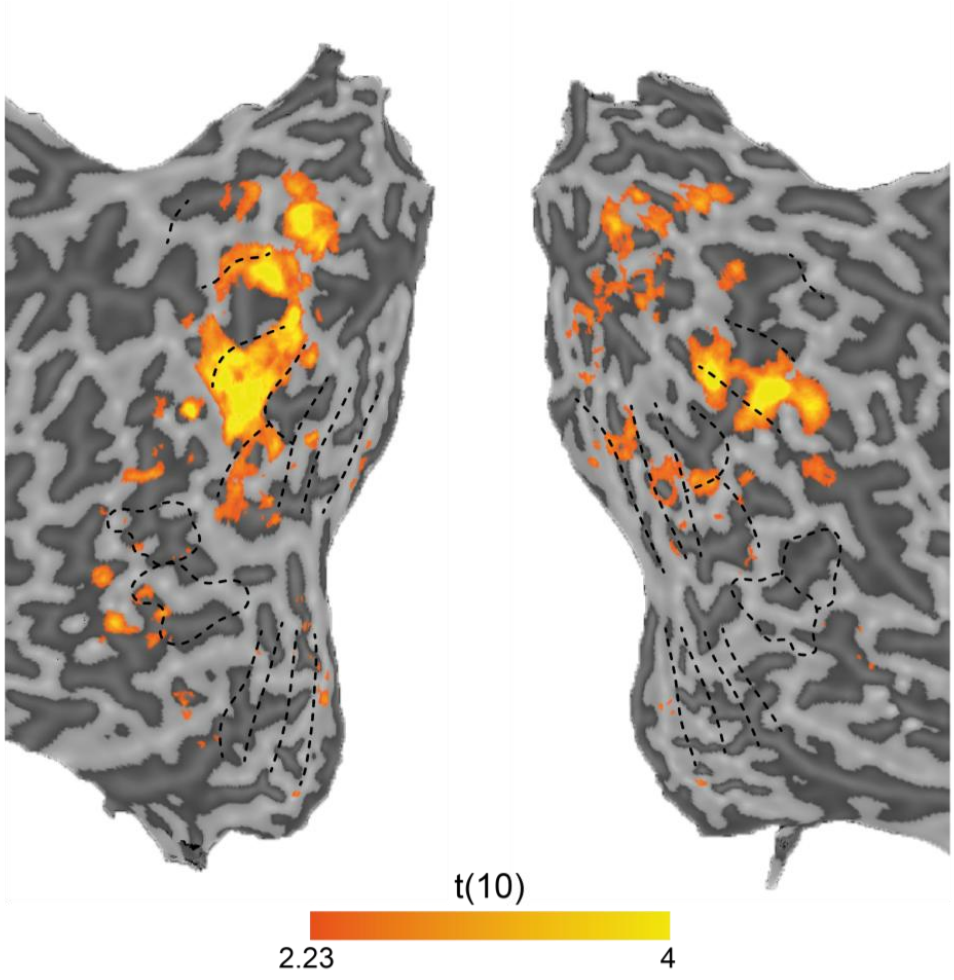


Figure 4.6: Between-subjects results of a searchlight classification presented on one individual's flatmap. All highlighted voxels are significant when contrasted to chance performance (t-test, $p < 0.05$). Boundaries of visual areas are identified through dotted lines.

the perceptual interpretation required that participants maintain task engagement during the delay period, minimising the potential for differences in attentional allocation for the different signal levels. Therefore, our findings appear more compatible with the perceptual interpretation of disparity information rather than general attentional processes.

4.4 Discussion

This study examined the relationship between perceptual judgments and cortical activity in the human visual cortex. Specifically, we manipulated the proportion of anticorrelated dots

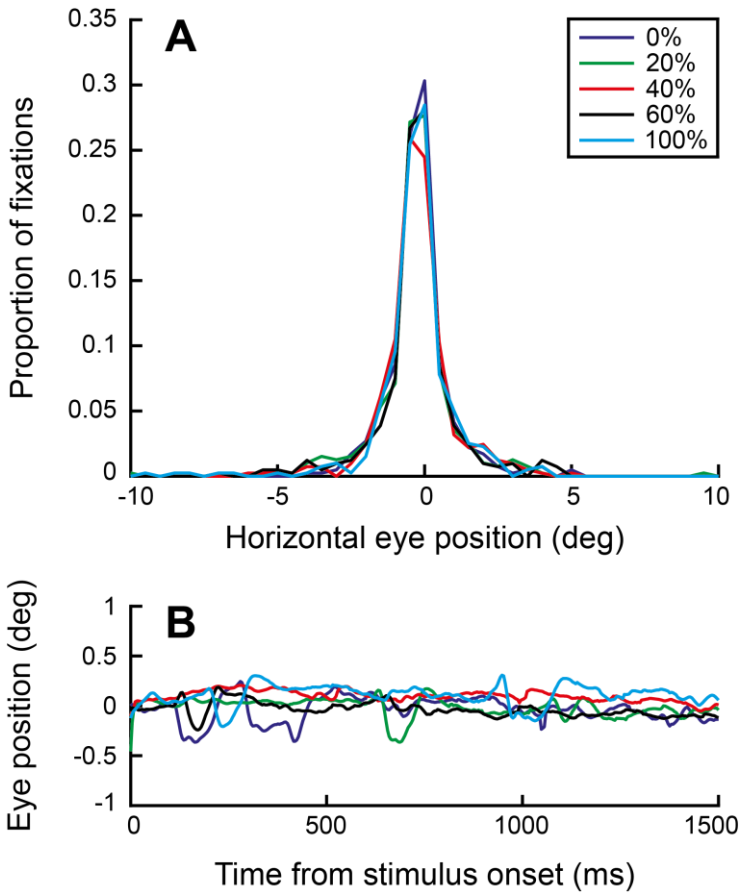


Figure 4.7: Eye-movement analysis. (A) Horizontal eye movements of three participants that were measured during completion of the behavioural task. This graph shows the proportion of time in each horizontal position where a fixation measurement was recorded, where 0 degrees represents the point of fixation. (B) Average horizontal eye position from the time of stimulus onset.

presented to observers and identified regions that reflected the depth percept of the observer and had therefore solved the stereo correspondence problem. Our results advance the understanding the neural representation of binocular depth perception by revealing that for the dorsal stream, the stereo correspondence problem is likely solved by later dorsal regions as both V7 and parietal region VIPS were highly diagnostic of perceptually useful, disparity-defined depth.

4.4.1 Disparity processing in the parietal cortex

Previous investigations have shown that activity in V1 does not reflect depth perception (Cumming & Parker, 1997, 2000) and, in contrast, activity in the anterior parietal region AIP shows an attenuated response to anticorrelated (but not correlated) stimuli (Theys et al., 2012b), thereby matching the percept of the observer. Here, our results indicate that these boundaries can be more conveniently reduced for the dorsal stream where the correspondence problem is solved and the depth percept realised by higher dorsal areas V7 and VIPS.

Further, our finding that V7 and VIPS reflects the perceptual interpretation of disparity signals is in agreement with the role of prior and subsequent dorsal regions in disparity processing for the dorsal visual pathway. Neuroimaging evidence has supported the role of both V7 and VIPS in disparity processing (Durand et al., 2009; Georgieva et al., 2009; Minini et al., 2010; Preston et al., 2008), though the loci of disparity processing is often considered to be located in intermediate dorsal region V3A (Backus et al., 2001; Cottereau et al., 2011, 2012a; Tsao et al., 2003). It is therefore possible that responses in these earlier dorsal visual areas compute information regarding disparity and then combine this information in later dorsal and parietal regions. Particularly, previous studies have supported combining depth estimates in parietal regions from both binocular disparity (Sakata et al.,

1998; Shikata et al., 1996; Taira et al., 2000; Tsutsui et al., 2001; Tsutsui et al., 2002) and monocular cues (Orban et al., 2006a; Tsutsui et al., 2001; Tsutsui et al., 2002; Vanduffel et al., 2002) to indicate a more generalised representation of depth (Sakata, Tsutsui, & Taira, 2005; Tsutsui, Taira, & Sakata, 2005). Although a generalised representation does not necessarily equate to depth perception, it indicates a global percept that is likely to have solved the correspondence problem. Further, posterior parietal regions are likely to have access to the depth percept as neuronal activity in the monkey lateral intraparietal area (LIP) has been related to its perception of motion (Shadlen & Newsome, 1996; Williams et al., 2003). The literature therefore supports our finding that responses in later dorsal regions mirror the perceptual judgments of the observer, which can only occur after stereo correspondence has taken place.

4.4.2 Comparing the fMR-metric functions to previous findings

Preston and colleagues (2008) previously measured classifier performance to correlated and anticorrelated stereograms in the human brain, and found a preference to correlated RDS in all of their measured dorsal regions and in ventral region LO. Our study extends on these findings in two important ways. First, our study varied the proportion of anticorrelated dots present in the stimulus and examined intermediate values such that we could create a psychometric function that reflected visual perception. Rather than compare differences in activity between correlated and anticorrelated stimuli, we identified the perceptual interpretation and found activity in cortical regions that reflected this performance. Second, their study involved passive viewing where participants were not required to pay specific attention to the presented depth and instead performed an attentional control task at the fixation point. The near-far depth discrimination task performed by participants in our

experiment meant that our results included neural processes involved in the making of the perceptual decision.

It is therefore not surprising that results somewhat differ between these two studies. The most prominent difference is the lower prediction accuracies observed for this study (~65% in any visual area for this study in comparison to up to ~85% in Preston *et al.*, 2008). The purpose of using active viewing in this paradigm was to measure behavioural performance concurrent to fMRI responses. Thus our conditions were required to vary from trial-to-trial in an event-related design. However, results for these designs are known to produce noisy estimates due to the successive presentation of different conditions before the hemodynamic response has returned to baseline. While we took measures to reduce the impact of this noise, including counterbalancing conditions and combining the activation from multiple trials prior to classification, this design and its inherent noise reduced the sensitivity of the MVPA classifier. Thus, while Preston *et al.* (2008) found significant activation to correlated RDS in all of their measured cortical regions (and also for several visual areas in response to anticorrelated RDS), using an event-related design we observed significant activation in fewer visual areas. Specifically, unlike the previous authors we did not find significant prediction accuracies for V1, V3v, LO and hMT+/V5 (**Figure 4.4**). However, a result below the criterion for significance should not necessarily be taken to imply that the region is not involved in the processing of disparity signals, but rather that our data acquisition and/or analysis methods were insufficiently sensitive to detect contributions from these regions.

Likewise, when prediction accuracies are located close to chance, a fit of the psychophysical results does not imply a relationship to perception. Although by nature of the paradigm prediction accuracies for lower proportions of anticorrelated stimuli are expected to

be located at, or close to chance, it is possible random fluctuations can result in a significant fit of the fMR-metric function. We minimised this occurrence by only fitting psychometric functions to ROIs that performed significantly above chance for perfectly correlated RDS. However, it is difficult to separate random fluctuations around chance from results that were simply less sensitive, such as in V3B/KO (**Figure 4.5**).

Preston and colleagues (2008) also found a significant difference in the decoding ability of the classifier between correlated and anticorrelated RDS for ventral region LO and in all of their measured dorsal (V3d, V3A, V3B/KO, V7, hMT+/V5) and parietal (VIPS, POIPS, DIPSM) regions. Here, the depth judgments of our observers corresponded to changes in the prediction accuracy for later dorsal regions V7 and VIPS, with suggestive evidence towards this relationship found in earlier dorsal regions V3A and V3B/KO. This potential role of earlier dorsal regions is consistent with previous studies that have shown responses to the global properties of stimuli as early as V3d (Bridge & Parker, 2007; Chandrasekaran et al., 2007; Georgieva et al., 2009; Preston et al., 2008) and V3A (Backus et al., 2001; Tsao et al., 2003), though further investigation is required to identify the role of these regions in stereo correspondence.

Interestingly, our results did not reveal sensitivity to the perceptual interpretation in any ventral visual area. This is at odds with previous studies that have observed an attenuation of responses to anticorrelated stimuli in macaque area V4 (Kumano et al., 2008; Tanabe et al., 2004) and an absence of responses to such stimuli in the inferior temporal cortex (Janssen et al., 2003). Further, fMRI studies have shown activity in relation to the binocular percept in LO (Chandrasekaran et al., 2007; Preston et al., 2008). It therefore seems likely that our paradigm was not sensitive enough to measure responses in ventral visual areas. Moreover, our planar stimuli may not adequately stimulate the neuronal populations in ventral regions,

as it is possible that these may prefer more complex shape elements (Janssen, Vogels, & Orban, 1999; Janssen et al., 2000a; Orban, 2011) though these regions have been previously found to respond to frontoparallel surfaces (Hegde & Van Essen, 2005; Preston et al., 2008). Alternatively, the difference in responses between visual pathways may have been a result of our stimulus favouring the type of processing that occurs in dorsal visual areas. For example, ventral regions may use a sparser code than other areas, resulting in less biased voxel responses and thereby reducing the performance of the MVPA classifier.

4.4.3 Comparison to signal-in-noise and feature difference tasks

The results of this experiment were largely consistent with what we observed in both the signal-in-noise and feature difference tasks in Chapter 3. Specifically, in both cases we observed perceptually-relevant responses in later dorsal visual areas. However, we observed a response in earlier regions for this experiment (V7, VIPS) than we did for the coarse and fine disparity tasks (VIPS, POIPS). The most likely explanation for this difference is the use of anticorrelated dots for this experiment. Only regions that are involved in local matching will respond to anticorrelated RDS, and once the correspondence problem has been solved the stimulus no longer yields a depth percept or elicits a neural response (Janssen et al., 2003; Theys et al., 2012b). In contrast, solving the correspondence problem does not aid in discrimination of the target depth for either the signal-in-noise or feature difference tasks. Instead, the difficulty arises from detecting signal from noise where perception is limited by external and internal interference for the signal-in-noise and feature difference tasks respectively. This necessitates an additional step of processing after stereo correspondence, and it is therefore not surprising that perceptual activity is observed in a later visual region. Further, the use of anticorrelated dots restricts the presentation of dots to a single disparity

magnitude across all conditions, such that the behavioural percept is affected but the overall stimulus properties remain constant. Similarly, we suggest that we did not observe parametric variation of classification performance in line with depth judgments for POIPS as it may have been involved in this additional stage of discriminating signal-from-noise.

In summary, we manipulated behavioural performance by varying the proportion of anticorrelated dots in a random-dot stereogram while simultaneously measuring fMRI activity. We used a multivariate classifier to identify cortical regions that had solved the correspondence problem and were involved in the perception of disparity-defined depth. We found activity in later dorsal regions V7 and VIPS mirrored stereoscopic depth judgments and concluded that for the dorsal visual pathway, stereo-matching is performed prior to the posterior parietal cortex.

5. Depth configurations

Abstract

Binocular disparities evoke widespread fMRI responses in the visual cortex, but responses appear to differ depending on the type of configuration used. To understand the impact that configuration has on the cortical processing of disparity, we presented participants with random-dot stereograms in a centre-surround design and measured cortical activity throughout the visual cortex. We independently varied the disparities of both centre and surround and examined cortical responses when we (1) varied the relationship between centre and surround, or (2) shifted the depth position of the entire configuration. We developed three tests using a multivariate classifier that discriminated between depth configurations. First, we examined the effect of changing the surround location when the target disparity was kept at a fixed depth. Second, we investigated the impact of increasing the disparity difference between two configurations. Third, we trained the classifier to discriminate disparity sign using one configuration, and tested the classifier's predictions for depths indicated by a new configuration. We consistently found that early visual areas, particularly V2, responded to changes in the disparity edge; intermediate dorsal regions (V3A, V3B/KO, V7) were able to extract depth information using either fixation or the surround as a reference, while responses in ventral regions (V4, LO) preferentially calculated disparity with reference to its surround.

5.1 Introduction

Binocular disparity results from the horizontal separation of the eyes, where slight differences in the images presented to each eye can give rise to a perception of depth (Wheatstone, 1838). The human brain is exquisitely sensitive to these differences, and they are involved in a multitude of visual tasks [e.g., figure-ground segregation (Qiu & von der Heydt, 2005; von der Heydt et al., 2000), 3D shape processing (Chandrasekaran et al., 2007; Georgieva et al., 2009), grasping (Culham et al., 2003), and making vergence eye movements (Masson et al., 1997; Takemura et al., 2001)]. Given the role of disparity in a range of functional tasks, it is perhaps not surprising that binocular disparities have been observed to evoke widespread fMRI responses in the human visual cortex (Backus et al., 2001; Georgieva et al., 2009; Minini et al., 2010; Neri et al., 2004; Parker, 2007; Preston et al., 2008; Tsao et al., 2003).

This widespread cortical response is also reflective of the numerous computations that are required for binocular depth perception, ranging from the pairing of images from the two eyes to forming a representation for a complex 3D surface. As such, visual areas are likely to favour a specific type of encoding that reflects the computational task for which it processes. For example, neurons in the inferotemporal cortex of the ventral visual pathway have been observed to respond to convex or concave surfaces in line with its role in 3D shape processing, but this response was frequently inhibited when the surface was approximated by either slanted or frontoparallel planes (Janssen et al., 2000b). Specificity to the depth order of the disparity stimulus has been observed in several cortical locations, including V4 (Hegde & Van Essen, 2005), TEs (Janssen et al., 2000b; Joly, Vanduffel, & Orban, 2009) and CIP (Taira et al., 2000). Furthermore, results may be characteristic to only a particular type of configuration. Relative disparity selectivity was not observed in MT neurons when using a centre-surround stimulus (Uka & DeAngelis, 2006), but a higher level of selectivity was

observed in this same region when using transparent superimposed planes (Krug & Parker, 2011). Similarly, the introduction of a gap between a centre and surround stimulus reduced the EEG response in only some of their measured regions, revealing that different cortical regions respond to changes in the configuration differently (Cottereau, McKee, & Norcia, 2012b).

It remains an open challenge to understand the impact different configurations have on the processing of disparity. Here, we measured the fMRI response across visual areas while observers viewed depth configurations (i.e., stimuli consisting of test planes at different depths) comprised of random dot stereograms in a centre-surround configuration. We investigated the importance of using either fixation or the surround as a reference in disparity calculations. This distinction is important as a preference towards fixation is indicative of elementary processing that is based directly on retinal inputs, while a preference towards using the surround as a reference suggests more abstracted representations that show tolerance to manipulations of the stimulus. Neurophysiology studies have compared responses with reference to fixation and the surround through examination of absolute and relative disparities (e.g., Cumming & Parker, 1999; Thomas et al., 2002); however, it is difficult to separate these properties in the scanner environment. We therefore performed two manipulations to investigate the cortical areas whose activity was related to the processing of depth configurations. First, we varied the relationship between centre and surround disparities to identify the dependence on processing disparity relative to fixation (**Figure 5.1a**). Second, we shifted the position in-depth of the entire stimulus to identify areas that used disparity information from the surround (**Figure 5.1b**). We used multivariate classification software to classify between depth configurations for each participant and developed analyses to examine the effect of the configuration on visual areas. We found that early visual area V2 responded

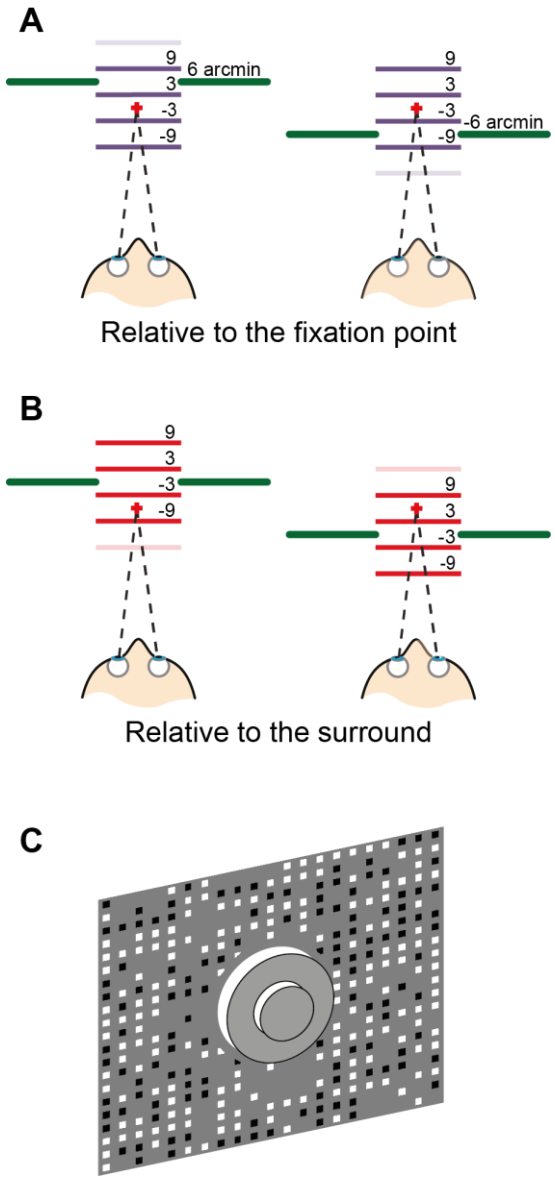


Figure 5.1: Stimulus conditions and illustration. (A) An aerial view of the eight stimulus conditions used for analyses considered relative to the fixation point. For each configuration, one central (purple) plane is presented with the corresponding surround (green) plane. Each plane is labelled with the depth between the central test plane and fixation, in arcmin. Dimmed planes are the conditions that did not have an appropriate match at the other surround location and were not used for this set of analyses. (B) The respective aerial view of the eight stimulus conditions used for analyses considered relative to the surround. Each plane is labelled with the depth between central and surround planes, in arcmin. (C) An illustration of the stimulus viewed from the front.

to changes to the disparity edge; responses in dorsal regions (V3A, V3B/KO, V7) were able to make use of information from both fixation and the surround; while ventral regions (V4, LO) preferentially responded to disparities with reference to its surround.

5.2 Methods

5.2.1 Participants

Eight participants (four females) were recruited from the University of Birmingham, including author MLP. Mean age was 26.1 years (range, 21 – 33 years). All participants had normal or corrected-to-normal vision with no deficits in colour vision and were paid for their participation. Prior to scanning, all participants were screened for stereo deficits in the laboratory haploscope to ensure stimuli presented in the study were above observers' detection thresholds. The experiment was granted ethics approval and all participants provided written informed consent.

5.2.2 Stimuli

We used correlated random dot stereograms (RDS) consisting of black and white dots and displayed on a mid-grey background. Stimuli consisted of a circular central plane (7° diameter) and an annular surround (14° diameter) which was used as a pedestal disparity (**Figure 5.1c**). When the surround was an uncrossed disparity (6 arcmin), the central patch was one of five disparities (-9, -3, 3, 9 or 15 arcmin). Similarly, when the surround was a crossed disparity (-6 arcmin), the central patch contained similar disparities (-15, -9, -3, 3 or 9 arcmin), with only a change of sign to the 15 arcmin condition. The depths of the central plane were chosen so that four disparities of the same magnitude existed (±3, ±9 arcmin) when compared to either the plane of fixation (**Figure 5.1a**) or to the pedestal surround (**Figure 5.1b**). The dot density of the stereogram was 20 dots/deg^2 and each dot had its edges softened using a Gaussian distribution (diameter 0.15° at half-height). The background surrounding the RDS was filled with a grid of black and white squares which could be used as a clear reference and encouraged vergence to the plane of the screen. The fixation marker was

a hollow white square (0.55° side length) that was presented in the centre of the stereogram. Vertical and horizontal nonius lines (length 0.375°) were attached to the fixation marker to promote vergence at the plane of the screen. We restricted presentation of dots to outside of a circular region (1.5° diameter) centred at fixation to ensure the stimulus did not interfere with binocular fusion.

5.2.3 Design

Each run was comprised of ten different stimulus conditions corresponding to the ten different depth configurations. Conditions were presented in 16 second blocks, with 16 stimulus presentations each lasting 500 ms (inter-trial intervals of 500 ms). Blocks were repeated three times per run, with the ordering of conditions randomised across runs and participants. Each experimental run consisted of 3 repetitions of each stimulus block. Each scan lasted 512 s and included two 16 s fixation periods (one in the beginning and one at the end of the run). Fixation periods consisted of only the fixation square and the background squares being visible, with no random-dot stereogram presented. Participants completed approximately eight runs in a single session. During all runs, observers performed a nonius alignment test (Popple et al., 1998). A vertical line (length 0.22°) was presented inside the fixation square in one eye on 25% of trials slightly offset from centre. Participants were asked to report its horizontal position ('left' or 'right') with respect to the nonius line presented to the other eye. Participants viewed additional stimuli (20 trials per condition) prior to the commencement of scanning and were asked to judge the depth position of the central plane relative to its surround. Participants generally performed with a high level of accuracy (mean, 91.2%) and could adequately perceive the relative depth in these configurations.

5.2.4 fMRI data acquisition

The study was performed in a 3-Tesla Philips Achieva MRI scanner at the Birmingham University Imaging Centre. We used an eight-channel SENSE head coil to collect echo-planar imaging (EPI) and T1-weighted anatomical (1 × 1 × 1 mm) data. For experimental runs, EPI data [echo time (TE), 34 ms; repetition time (TR), 2000 ms] were obtained from 28 slices (voxel size, 1.5 × 1.5 × 2 mm, 256 volumes) positioned close to the coronal plane. Localisers were obtained in a separate session using the same EPI sequence. We used a pair of video projectors (JVC D-ILA SX21) to display stereoscopic images, with each projector fitted with a unique interference filter (INFITEC, GmBH) that distributed the wavelengths of visible light between the two projectors. The images from each projector were optically combined using a beam-splitter cube and passed through a waveguide into the scanner room. Stimuli were projected onto a translucent plastic screen located behind the head coil and inside the bore of the scanner. Participants viewed the screen via a mirror positioned on the head coil angled at 45°, with an optical path length of 65 cm. Since the INFITEC filters produced negligible overlap between the spectra for each projector, there was minimal crosstalk between the projected images for a participant wearing a pair of corresponding filters. Unique stimuli were pre-generated for each participant using C#, and the experiment was performed using MATLAB (The MathWorks, Natick, MA) and the PsychToolBox 3 extension (Brainard, 1997; Pelli, 1997).

For each participant, we identified regions of interest (ROIs) from independent data in a separate localiser scan. We used a rotating wedge stimulus and expanding concentric rings to define the borders of early retinotopic areas (V1, V2, V3v, V4, V3d, V3A and V7) (Aguirre et al., 1998; DeYoe et al., 1996; Sereno et al., 1995). In particular, V4 was defined as the ventral region adjacent and inferior to V3v containing a full hemifield representation (Tootell

& Hadjikhani, 2001; Tyler et al., 2005) and V7 was defined as the dorsal region adjacent and anterior to V3A, also containing a full hemifield representation (Tootell et al., 1998; Tsao et al., 2003; Tyler et al., 2005). We identified higher dorsal regions [V3B/kinetic occipital area (KO), human motion complex (hMT+/V5)] and the ventral lateral occipital region (LO) from additional localiser scans. Area V3B/KO (Dupont et al., 1997; Zeki et al., 2003) was defined anatomically as the dorsal region adjacent and lateral to V3A with which it shared a foveal representation (Tyler et al., 2005), and functionally as the set of voxels that responded significantly more ($p < .001$) to kinetic boundaries than to the transparent motion of black and white dots without clear borders. Area hMT+/V5 was defined as the region in the lateral temporal cortex that responded significantly more ($p < .001$) to an array of coherently moving dots than to an array of static dots (Zeki et al., 1991). The lateral occipital (LO) area was identified as the region in the lateral occipito-temporal cortex that responded significantly more ($p < .001$) to images of objects and shapes than to scrambled versions of these images (Kourtzi & Kanwisher, 2000, 2001). During all localiser scans, participants performed an attentionally demanding task on the fixation point, except for the LO localiser in which they had to respond if the same image was presented consecutively. For illustrative purposes, the mapping of all these ROIs for one participant is presented in **Figure 5.2**.

5.2.5 fMRI data analysis

We processed MRI data using BrainVoyager QX (BrainInnovation, Maastricht, the Netherlands). For each participant, anatomical scans were transformed into Talairach space and used for 3D cortex reconstruction, inflation, flattening and the segmentation of gray and white matter. Preprocessing of functional data included head movement correction (translation and rotation), slice scan time correction, and removal of low-frequency drifts and

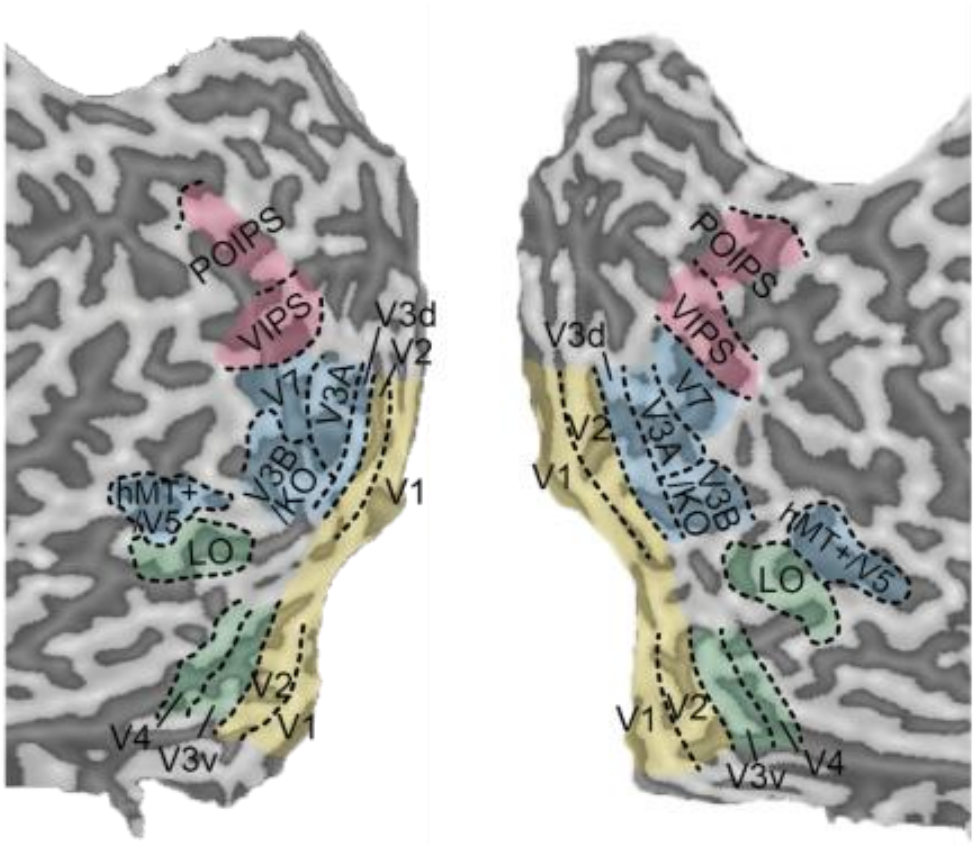


Figure 5.2: Regions of interest in one participant showing retinotopic areas, V3B/KO, hMT+/V5 and LO.

linear trends through temporal high-pass filtering (3 cycles). Functional runs were aligned to the participant's anatomical data and then transformed into Talairach space. Consecutive volumes that contained head movement greater than 1 mm of translation or 1° of rotation were excluded from further analysis.

We used Multi-Voxel Pattern Analysis (MVPA) software to analyse the fMRI data from each ROI. For voxel selection, gray matter voxels were isolated through a grey-matter mask from the anatomical scan and only included voxels that showed a *t* value larger than 0 when contrasting fMRI activity from all stimulus conditions to the fixation baseline across all experimental runs. These were then sorted according to their response (*t* statistic) and the 300 most active voxels (or highest number available) in each ROI were used in the classification.

Each block of functional data was averaged to generate a single training pattern. The time series for each voxel was then normalised (*z*-score) in each experimental run to compensate for the decrease in mean image intensity with distance from the receiver coil. The fMRI time series was shifted by 2 volumes (4 s) to account for the hemodynamic delay of the BOLD signal. Each volume had the mean univariate signal subtracted to remove any bias created from it.

We used a leave-one-run-out cross-validation sampling procedure, to ensure generalisation of the classification. For each cross-validation, one run was left out as an independent test dataset and the data from the rest of the runs was used as the training set. The classification accuracy for each ROI was obtained by averaging the prediction accuracy across cross-validations. Training and test patterns were calculated for all depth configurations independently. The reported prediction accuracy of the classifier corresponds to the proportion of trials on which it could correctly predict the stimulus based on the pattern of the fMRI response. In our two-way classifications, a prediction accuracy of 0.5 reveals performance based on chance and a prediction accuracy of 1 reveals perfect discrimination between the two stimulus categories. For each ROI, prediction accuracies were averaged across participants.

Fixed depth. In the first of our analyses, we examined whether the classifier could discriminate a change in the surround location when the target disparity was held constant at 9 arcmin (results were combined for both crossed and uncrossed disparities). We compared prediction accuracies when disparity was measured with respect to either fixation (e.g., **Figure 5.3a**) or the surround (e.g., **Figure 5.3b**).

Function of step size. We investigated the effect of disparity magnitude across visual areas. We performed a two-way classification on the disparity sign (near *vs.* far), and

examined decoding ability for disparities of varying sizes (6, 12, 18, 24 arcmin) (e.g., **Figure 5.4a**). This was repeated in both crossed and uncrossed surround locations, with the resulting prediction accuracies averaged together. We performed this twice; once with disparities relative to fixation and once when they were relative to the surround. For the 6 and 18 arcmin disparity conditions, planes were equidistant from the reference. However, for the 12 arcmin condition, we selected target disparities that did not encompass both fixation and the surround, though this was consequently not-centred (-9 and +3 arcmin, or with opposite signs). Similarly, since the 24 arcmin encompassed both fixation and the surround, results for this were common between both analyses.

Transfer Test. To obtain a test of specificity in the processing of depth configurations, we asked whether information from one depth configuration is diagnostic of depth indicated by a different configuration. We performed a transfer test whereby we trained the classifier to discriminate disparity sign (±9 arcmin) using one configuration, and tested the classifier's predictions for data obtained when depth was indicated by a new surround location. The depth of the target plane, across transfer, was kept constant relative to either fixation (e.g., **Figure 5.6a**) of the surround (e.g., **Figure 5.6b**). In each case, the result was combined with a transfer performed in the other direction. To assess the relationship between transfer classification performance (d'_T) and classification performance without transfer taking place (d'_{NT}), we calculated a bootstrapped transfer index,

$$I_T = \frac{d'_T}{d'_{NT}} \ .$$

To assess the amount of transfer that arises by chance and form a baseline to judge significance of our observed results, we conducted the transfer test on randomly permuted

data (i.e., we randomised the correspondence between fMRI data and training labels and estimated the classifier prediction for each visual area) with 999 shuffles per area. The shuffled prediction accuracies were averaged and its sensitivity (d'_{Sh}) was used to calculate the transfer index of shuffled responses,

$$I_{Sh} = \frac{d'_{Sh}}{d'_{NT}} \ .$$

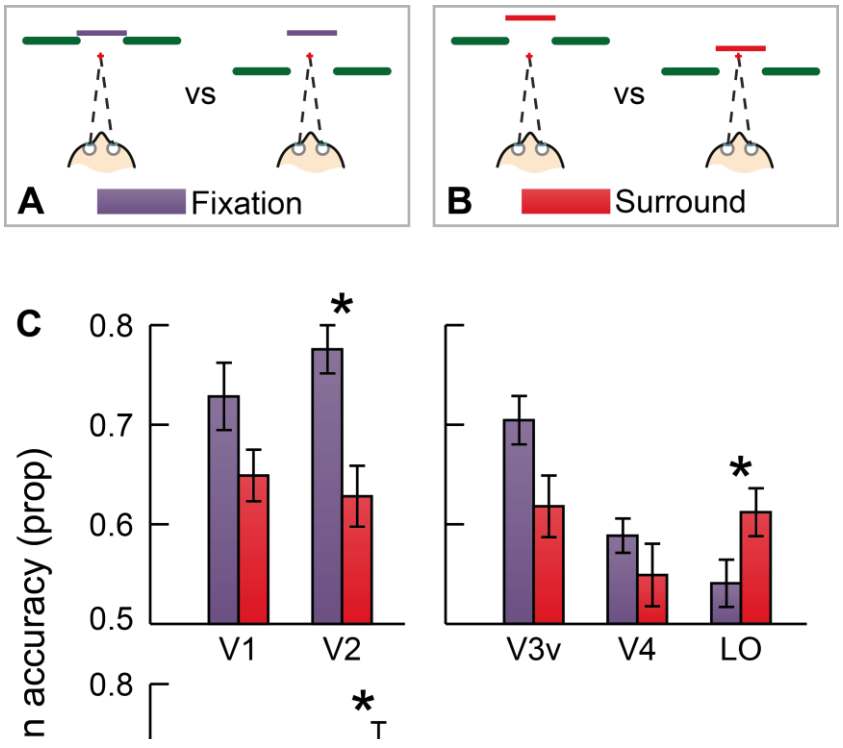
5.3 Results

We presented participants with random-dot stereograms that depicted a centre-surround configuration of two depth planes while we measured fMRI responses in regions of interest across the visual cortex. To test the importance of the configuration of the planes, we manipulated the disparities of both planes independently and compared the response to configurations where the central disparity was kept constant relative to fixation, and configurations where the disparity difference between centre and surround was held constant. We examined responses in early visual areas (V1, V2), ventral regions (V3v, V4, LO) and dorsal regions (V3d, V3A, V3B/KO, V7, hMT+/V5) that have been previously implicated in the processing of disparity information in the human brain (Backus et al., 2001; Minini et al., 2010; Neri et al., 2004; Rutschmann & Greenlee, 2004). Having extracted fMRI signals, we trained a linear support vector machine to associate patterns of voxel activity within each ROI to the depth configuration that gave rise to the activity. We tested whether we could predict the viewed stimulus from the fMRI activity, calculating the mean leave-one-run-out prediction accuracy for a classifier trained to discriminate between different depth configurations. We predicted that changes to the depth configuration would elicit different

responses depending on the type of disparity processing that is taking place within each visual area.

5.3.1 Comparing classifier performance for centre-surround configurations that contained a fixed depth relative to fixation or the surround.

We first compared decoding accuracies between stimulus configurations when they contained



	$t_{(7)}$	*p*-value
V1	1.939	.094
V2	4.345	**.003***
V3v	1.952	.092
V4	0.907	.395
LO	-3.627	**.008***
V3d	2.064	.078
V3A	-2.746	**.029***
V3B/KO	-0.901	.397
V7	-2.707	**.030***
hMT+/V5	-0.419	.688

Table 5.1: Two-tailed t-test for each ROI in the fixed depth analysis. An asterisk marks a significant fit for that ROI ($p < .05$).

a fixed depth relative to the fixation point or the surround. In particular, we examined the difference between using fixation as a reference in contrast to the surrounding pedestal disparity. For this analysis, above-chance prediction accuracies represent a change between the two configurations that the classifier has been able to identify and exploit. A two-way ANOVA revealed that the classifier could extract diagnostic information about disparity in the measured visual areas ($F_{(9,63)} = 10.91$; $p < 0.001$). However, in both cases, the change in stimulus configuration accompanied a change in the depth of the surround. Thus, significant decoding accuracies could simply be due to the change in the depth of the surround. However, as the same change in the surround occurred for both configurations relative to fixation and the surround, we examined the difference in the prediction accuracies between these two configurations. Comparing prediction accuracies for disparities relative to fixation and the surround showed a significant interaction between the stimulus configuration used and the cortical region of interest ($F_{(9,63)} = 5.70$; $p < 0.001$). In particular, early visual areas (V1, V2) and V3 (both dorsal and ventral portions) had larger classification accuracies for configurations defined relative to fixation, while later ventral (LO) and dorsal (V3A, V7) areas were the inverse, and had higher classification accuracies when the surround was used as a reference. In particular, examining these differences revealed that the classifier was able to decode stimulus configurations significantly in area V2 better when they were defined relative to the fixation point. In contrast, prediction accuracies were significantly higher in ventral LO and dorsal V3A and V7 when stimuli were contrasted relative to the surround (**Figure 5.3c**; **Table 5.1**). This suggests important transformations in the way in which disparities are represented between early and higher visual areas.

We investigated changes in the stimulus properties that might explain the difference in prediction accuracies between early and later visual areas. First, we notice for the analysis

where the disparity of the central plane remains constant relative to fixation (e.g., **Figure 5.3a**), there was a large change in the magnitude of the disparity edge, and the size of this disparity step appears to be particularly important for early visual areas. In line with our results, early visual area V2 has previously been implicated in the processing of location and orientation of disparity edges (Bredfeldt & Cumming, 2006; Qiu & von der Heydt, 2005; von der Heydt et al., 2000). We therefore suggest that it is this change in disparity edge that the classifier is using to discriminate responses in this analysis. Second, note that when we held disparity constant relative to the surround while changing the disparity of the surround (e.g., **Figure 5.3b**), there was a change in the overall depth position of the stimulus. Here, the relationship between centre and surround and therefore the disparity edges remained the same. Instead, the classifier likely discriminated fMRI activity relating to the depth position of the configuration, in line with a global processing of the stimuli that is frequently observed in later visual areas (Parker, 2007). We similarly observed significantly higher prediction accuracies in both later ventral area LO as well as in dorsal regions V3A and V7, indicating a role of these areas in the processing of the global depth structure.

5.3.2 Examining prediction accuracies as a function of step size

To examine changes in fMRI activity for varying magnitudes, we compared prediction accuracies for binary classifications when the classifier was trained to discriminate depth configurations where the difference between near and far stimuli had different disparity magnitudes (steps of 6, 12, 18 and 24 arcmin). We found that step size affected classifier performance when the target planes were considered relative to fixation ($F_{(3,21)} = 10.77$; $p < 0.001$) as well as when they were considered relative to the surround ($F_{(3,21)} = 20.57$; $p < 0.001$). This suggested that changes in the magnitude of the disparity difference were related

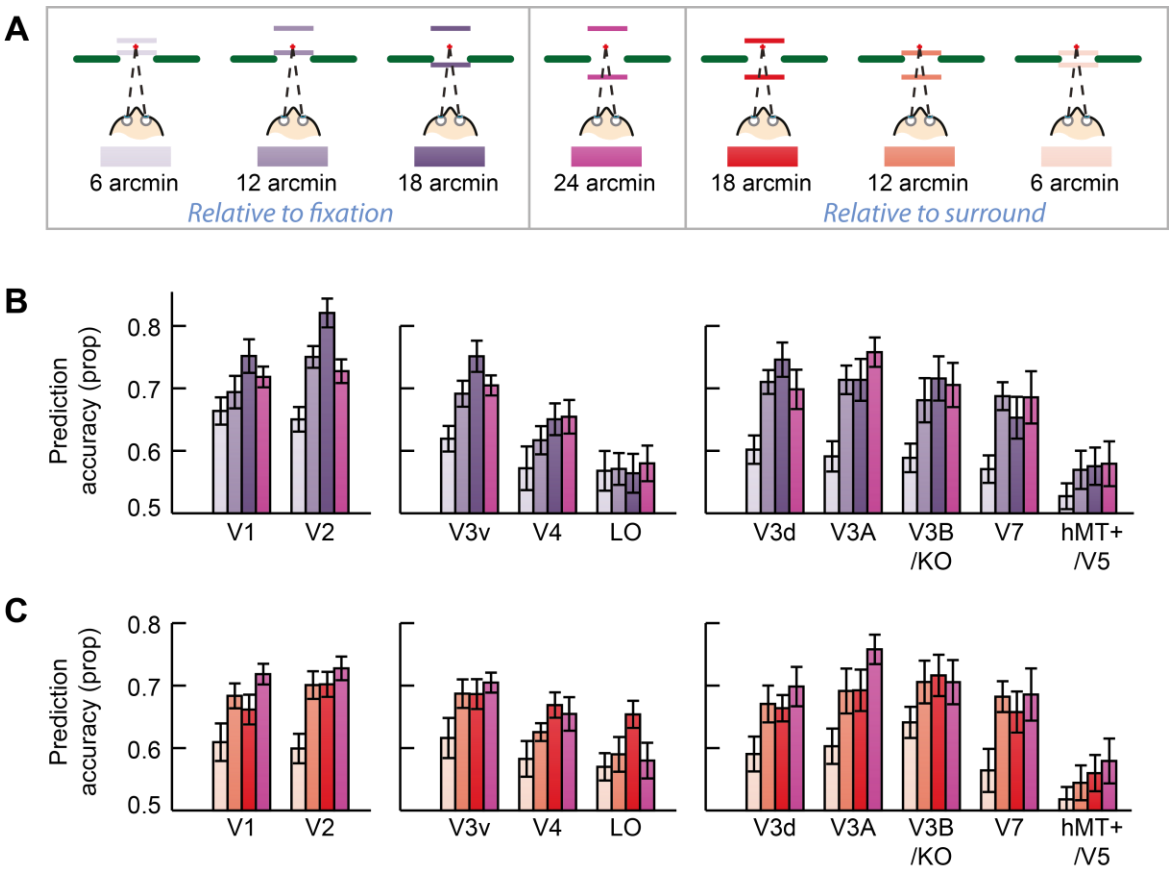


Figure 5.4: Illustration and results for the function of step size analysis. (A) An illustration showing conditions that were used for near-far classifications of different magnitude, where depth position was defined relative to fixation or the surround. The 24 arcmin (magenta) condition was common for both analyses. (B) Prediction accuracies in each ROI for different step sizes when compared relative to fixation. The disparity magnitude for each bar colour is labelled in (A). (C) Prediction accuracies in each ROI for different step sizes when compared relative to the surround. Error bars are standard error.

to changes in prediction accuracy. Generally, larger step sizes corresponded to larger prediction accuracies across the majority of visual areas. Further, we observed an interaction between visual areas and the magnitude of the disparity difference when relative to fixation ($F_{(27,189)} = 1.85$; $p = 0.01$), but not when relative to the surround ($F_{(27,189)} = 1.02$; $p > 0.05$).

In the majority of visual areas, classification performance generally increased as the magnitude of the disparity difference between the two planes increased (**Figure 5.4b–c**). However, we observe a decrease in prediction accuracies at the largest step size of 24 arcmin when relative to fixation for early visual areas as well as the ventral and dorsal components of

V3, respectively (**Figure 5.4b**). This suggests that activity in these areas may be involved in the processing of disparity edges. For step sizes smaller than 24 arcmin, classifications involved one configuration close to the surround (with a small disparity edge) while the other configuration was further away from the surround (with a large disparity edge). The step size of 24 arcmin is the only discrimination where both configurations are presented away from the surround. contain large (but unequal) disparity edges. Therefore, the small decrease in prediction accuracy in these visual areas likely relates to the increased similarity of the disparity edge between the two configurations.

To identify differences between **Figure 5.4b** and **Figure 5.4c**, we examined the difference in prediction accuracies depending on the reference used (**Figure 5.5**). Across step sizes, we observed higher prediction accuracies when relative to fixation for early visual areas V1 and V2, as well as for the ventral and dorsal parts of V3. This difference is also likely to reflect the processing of disparity edges, since configurations relative to fixation always contained unequal differences in disparity edge. Further, we observe minimal changes to prediction accuracies regardless of the reference used in dorsal visual areas as well as for

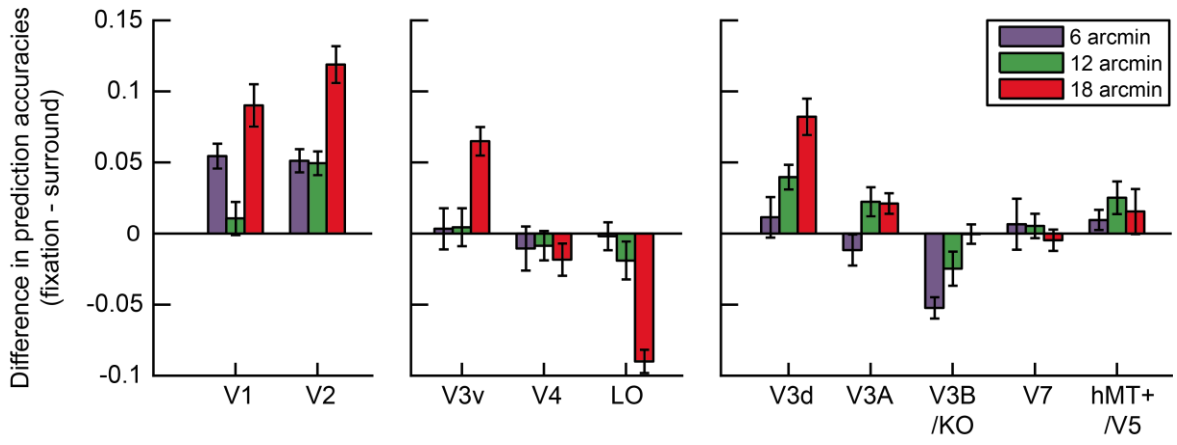


Figure 5.5: Difference in prediction accuracies in the function of step size analysis. Each bar represents the change in prediction accuracies between Figure 5.4b and Figure 5.4c (relative to fixation – relative to surround). Error bars are standard error. The 24 arcmin condition is excluded as it was the same for both analyses.

intermediate ventral region V4. This suggests similarity in the disparity processing when using surround or fixation as the reference. It was only ventral region LO that showed increased sensitivity to classifications that used the surround as a reference.

Our results appear to diminish in ventral visual areas (**Figure 5.4b–c**), to the point where responses in LO largely appear constant across most step sizes. It is possible that this type of response to varying magnitudes of disparity may reflect a processing that is invariant of disparity magnitude, and is instead only concerned with the sign of the disparity. This pattern of response would suggest a categorical processing of disparity, consistent with the response of neurons tuned to 'near' and 'far' disparities (Poggio, 1995).

5.3.3 Transfer test

To test for similarities between different depth configurations, we asked whether information from one configuration could indicate depth for a different configuration. In particular, we investigated whether learning from one classification would transfer to a new configuration when the surround was changed (e.g., **Figure 5.6a**) or when the entire configuration was shifted in-depth (e.g., **Figure 5.6b**). Our transfer index represented the ability of the classifier to transfer the learned pattern of fMRI responses from one configuration to that of another configuration. A value of 0 revealed an inability to transfer this information to a new configuration and a value of 1 indicated perfect transfer performance where responses to the transferred configuration were equivalent to performance without transfer. We assessed the amount of transfer that arises by chance by calculating the transfer index using randomly permuted data. We used the upper limit (two-tailed, 97.5%) of this shuffled distribution as a baseline for chance performance (**Figure 5.6c**, dashed lines).

Our results highlighted the role of intermediate dorsal regions in the encoding of depth signals. When the disparity of the surround was shifted to a new location, there was significant transfer performance in dorsal regions V3A and V3B/KO (**Figure 5.6c**, purple boxes). The results of the transfer with regard to fixation are broadly consistent with previous findings that observed a significant change in the EEG response for V3A and V4 ROIs when there was a change in the surround location (Cottereau et al., 2012a).

Furthermore, when we shifted the entire configuration in depth, we once again observed significant transfer in V3A and V3B/KO, as well as in intermediate ventral region V4 (**Figure 5.6c**, red boxes). Previously, an increasing number of neurons in the ventral stream of the monkey have been seen to respond similarly to depth configurations irrespective of their position in-depth (V2, Thomas et al., 2002; V4, Umeda et al., 2007). Our results mirror these neurophysiology findings, with our transfer index also increasing along the ventral stream up until region V4. However, results in the dorsal stream are more complex. Neurons did not respond in line with changes to the depth of the configuration in dorsal regions V3 and V3A (Anzai et al., 2011), or parietal region CIP (Taira et al., 2000), while results for MT have been mixed depending on the configuration used (Krug & Parker, 2011; Uka & DeAngelis, 2006). However, it is possible that these differences relate to changes in the stimulus. Neurophysiology studies typically present their stimulus in the periphery and the stimulus is tailored to suit the preferences of the individual neuron in disparity, motion and size. The lower sensitivity of our fMRI measurements means we were unable to isolate these properties and instead, our stimulus presentation encompasses neurons of varying preferences.

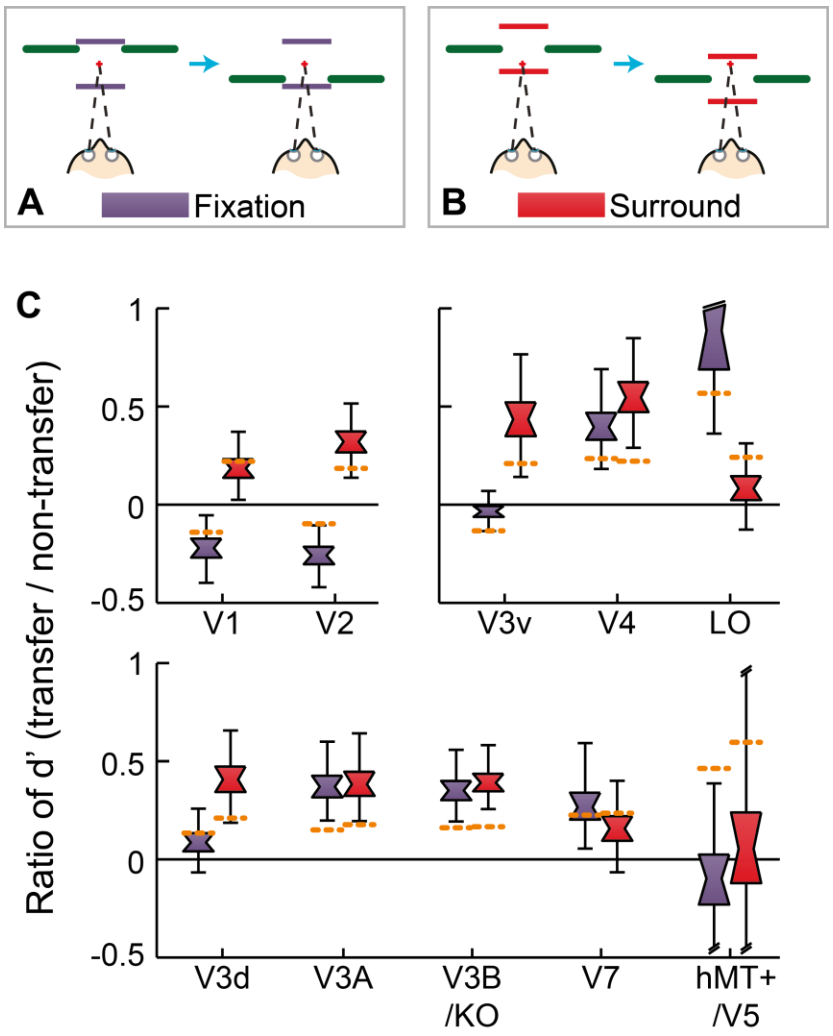


Figure 5.6: Illustration and results of the transfer test. (A & B) An illustration of the transfer task. The classifier was trained to discriminate between near and far on one pair of configurations, and then tested on a new pair of configurations. The depth remained constant between configuration pairs relative to fixation for (A) and relative to the surround for (B). (C) Results of the transfer task. The centre of each 'bowtie' is the median, the shaded area depicts 68% confidence values and the upper and lower error bars 95% confidence intervals. The orange dotted line is the baseline from the shuffled distribution.

In contrast, neuroimaging has revealed a similar transfer effect between motion and disparity in dorsal regions V3d and V3B/KO, which were thought to indicate a more general (integrated) representation of depth signals (Ban et al., 2012). Similarly, V3B/KO has been identified in the processing of depth structure, responding to binocular disparity even in the absence of subjective contours (Tyler, Likova, Kontsevich, & Wade, 2006). These results

suggest a more complex processing of disparity in intermediate dorsal regions, with these regions able to recruit information from multiple sources in their processing of disparity.

In addition, this analysis also supports our claim that the processing in early visual areas relates to the processing of the disparity edge. When the location of the surround was changed we observed significant transfer in V2, but interestingly this was in the opposite, negative direction (**Figure 5.6c,** purple boxes). This result suggests that the learning of depth sign (near or far) was reversed across transfer, corresponding to changes in the stimulus configuration. Since this was observed only when the depth was considered relative to the fixation marker and not when relative to the surround, we suggest that for early visual areas the classifier learns information about the magnitude of the disparity difference between the target and its surround. As illustrated in **Figure 5.6a**, the configuration of the far plane before transfer has a small disparity edge, while the configuration of the near plane contains a large disparity edge. Across transfer, this relationship is reversed and classification performance (and therefore our transfer index) occurs in the opposite direction. This supports our suggestion that early visual areas have not realised the depth percept and are performing only local processing, with classifier results being driven by the response to the disparity edges of the stimulus.

5.3.4 Responses in the visual parietal cortex

It should be noted that visual regions in the parietal cortex were localised and measured, but were subsequently excluded in order to simplify the presentation of results. Localisation of the ventral (VIPS) and parieto-occipital (POIPS) regions of the intraparietal sulcus was performed independently; defined as the set of voxels that responded significantly more ($p <$ .001) to 3D shape formed by disparity and structure-from-motion cues than random

disparities and motion speeds which did not form 3D shape (Chandrasekaran et al., 2007; Orban et al., 1999). All of our analyses were performed on these regions; however prediction accuracies were reasonably low across all analyses (0.63 or less), were considerably noisier for the transfer task (reminiscent of our results for hMT+/V5) and generally did not provide a clear result. However, we did observe significantly higher prediction accuracies for VIPS in the fixed depth analysis, when the disparity of the central plane remained constant relative to the surround in comparison to when it was defined relative to the fixation plane ($t_{(7)}$ = -4.221, p = .004). In contrast, there was no significant difference when we performed this same comparison in POIPS ($t_{(7)}$ = -1.112, p = .303). Our result for VIPS therefore mirrors our findings for other later visual areas (LO, V3A, V7) and suggests that VIPS is also involved in discrimination of the global depth position, though this information was not able to be transferred in depth.

5.3.5 Control analyses

We performed several control measures and designed our stimulus to ensure that the results were best explained through changes in the perceptual depth interpretation. First, we recorded horizontal eye-movements for three participants using a monocular limbus eye-tracker (CRS ltd). Analyses of these data suggested no systematic difference in eye-position (**Figure 5.7a–b**), and no statistical difference in the number of saccades between conditions, suggesting it was unlikely that eye-movements could adequately explain our results. The confines of the fMRI scanner and our use of spectral filters for stereoscopic presentation meant that we were unable to record changes in eye vergence; however we designed our stimuli to minimise this possibility. In particular, we instructed participants to maintain alignment of vertical and horizontal nonius lines that surrounded the fixation marker throughout all runs, encouraging vergence at the correct depth (the plane of the screen). Further, the stereogram was

surrounded by a clear reference pattern that was located at the fixation plane. Throughout all runs observers performed a subjective assessment of eye vergence (Popple et al., 1998). These results had minimal differences across conditions (**Figure 5.7c**), with a minor change in vergence for larger conditions. However, this change was not comparable to the change in depth of the stimulus, and suggests that vergence was relatively stable despite disparity manipulations. The use of a vernier task also meant that the motor response was decoupled from stimulus presentation, and attentional load was kept constant across all trials. Thus, it seems most plausible that our results are specifically related to the processing of depth configurations.

5.4 Discussion

To investigate the processing of disparity-defined depth configurations, we presented observers with random-dot stereograms in a centre-surround design and measured cortical activity throughout the visual cortex. We manipulated both centre and surround disparities and used a multivariate classifier to discriminate between particular depth configurations. We developed three tests to examine the effect of the configuration on visual areas. First, we tested whether the classifier could discriminate a change in the surround location when the target disparity was held constant. Second, we investigated the effect of increasing the disparity difference between two configurations. Finally, we trained the classifier to discriminate disparity sign using one configuration, and tested the classifier's predictions for depths indicated by a new surround location.

Results from all three analyses highlight the role of early visual areas in the processing of disparity edges, in line with a local processing of disparity. Prediction accuracies were significantly higher in V2 when the disparity of the target plane was kept constant but the

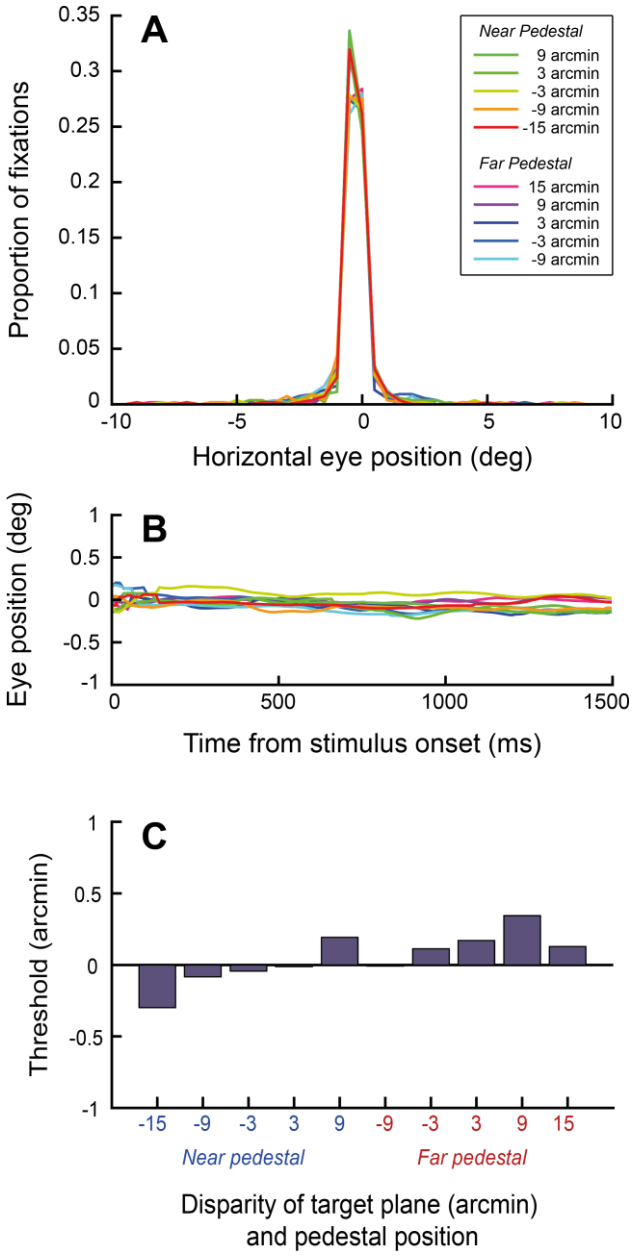


Figure 5.7: Eye-movement analysis and results of the vernier task. (A) Horizontal eye movements of three participants. This graph shows the proportion of time in each horizontal position where a fixation measurement was recorded, where 0 degrees represents the point of fixation. (B) Average horizontal eye position from the time of stimulus onset. (C) Mean threshold values for behavioural performance in the vernier task.

magnitude of the disparity edge differed, than when the disparity of the target plane changed and the disparity edge was of the same magnitude (**Figure 5.3c**). Similarly, classification performance in several early visual areas (V1, V2, V3v, V3d) decreased when there was increased similarity between the disparity edges in the two contrasted depth configurations (**Figure 5.4b**). Binary near-far classifications were consistently higher in these same visual areas when discriminating configurations with unequal disparity edges than when they were of equal but opposite magnitude (**Figure 5.5**). We also observed a significant negative transfer in V2 which corresponded to the magnitude of the disparity edge rather than the disparity magnitude of the target plane (**Figure 5.6c**, purple boxes). These results consistently reveal a role of early visual area V2 in the processing of disparity edges. Previously, responses in V2 have been seen to reflect the contours of stereoscopic perception and neurons fire in response to changes in the magnitude of the disparity edge, rather than the disparity of the surface itself (Bredfeldt & Cumming, 2006; Qiu & von der Heydt, 2005; von der Heydt et al., 2000). They are thought to undergo a transformation from the way V1 neurons are processed (Tanabe & Cumming, 2008) and can be explained as a feedforward model from the response of V1 neurons (Bredfeldt, Read, & Cumming, 2009). This is in agreement with our results, where we consistently observed a similar but nonsignificant trend for our results in V1 in these analyses. Similarly, it is suggested that the responses of V2 neurons are enough to represent the location and orientation of disparity edges without much further processing (von der Heydt et al., 2000) and supports processing of only local mechanisms (Qiu & von der Heydt, 2005). It should be noted that responses relating to disparity edges have been observed further down the visual pathway in V3A, where it may be used to support long-range perceptual judgments (Cottereau et al., 2012b). We did not find support for this idea, although it is possible this effect was masked in V3A by a greater response to changes in the overall

depth position (**Figure 5.3c**). These studies support our suggestion that the classifier is using local processing such as the disparity edge to discriminate changes in the stimulus.

Further, our study emphasises the role of dorsal visual areas in the processing of depth configurations. Classification performance in both V3A and V7 was significantly higher when the entire configuration was shifted in-depth compared to when it was only the surround that varied depth position (**Figure 5.3c**). Dorsal visual areas (V3A, V7, hMT+/V5) responded similarly to our classification for several magnitudes of disparity, regardless of whether disparity was measured from fixation or the surround (**Figure 5.5**). Intermediate dorsal visual areas (V3A, V3B/KO) were seen to transfer diagnostic depth information to a new configuration, regardless of whether the relationship between centre and surround was changed (**Figure 5.6c**, purple boxes), or the position in-depth was changed (**Figure 5.6c**, red boxes). Together, these results reveal the important role dorsal visual areas play in the processing of disparity-defined depth. Disparity stimuli when compared to equivalent stimuli presented at fixation are seen to evoke the strongest responses in dorsal regions, specifically V3A, V3B/KO and V7 (Backus et al., 2001; Tsao et al., 2003). It has also been shown that a ROI based at V3A responds to surround information in the processing of disparity, where a higher response was observed for a correlated than for an uncorrelated surround (Cottereau et al., 2011, 2012a). Our study supports these findings and suggests that intermediate dorsal regions use both fixation and the surround as a reference when representing disparity information.

In contrast, responses in the ventral stream appear to show a preference towards processing of disparities relative to their surround. We obtained higher responses in LO when the target disparity was kept constant relative to the surround than when the depth of the surround was changed (**Figure 5.3c**), revealing a preference for disparities to be processed

relative to the surround. Similarly, prediction accuracies in LO were higher when we performed a binary classification with reference to the surround (**Figure 5.5**). Finally, the index for our transfer task increased along regions of the ventral stream when the configuration was shifted in-depth, culminating at V4 which displayed significant transfer (**Figure 5.6c**). Interestingly, we did not observe this result in LO though neurons in later ventral region TEs responded to changes in depth for both planar surfaces (Janssen et al., 2000a) and 3D surfaces (Janssen et al., 2000b). Together, these results show an emphasis on disparity calculations being performed relative to the surround, consistent with the known role of the ventral stream in relative disparity processing (Neri et al., 2004; Shiozaki et al., 2012; Umeda et al., 2007). These are likely used to gather information about adjacent surface points and aid in the discrimination of 3D shape, consistent with the goals of the ventral visual pathway (Goodale & Milner, 1992; Janssen et al., 2000b; Joly et al., 2009; Parker, 2007). Although consistent with these findings, it is important to note that our study does not explicitly investigate relative disparity since both absolute and relative disparities were present in all of our conditions.

While we have made some clear advances in understanding the neural representations of depth configurations, we must be careful when interpreting these results. In particular, when we shift the depth of the entire configuration we are uncertain as to the underlying mechanism responsible for this process. It is possible that this reflects an invariance of depth position in the disparity processing for that visual area. In addition, it is possible that the results of our classifier have been influenced by the figure-ground organisation for different configurations. Specifically, responses in LO have been observed to be higher when a target disparity is presented in front of the surround (an object) compared to when it is presented behind the surround (a hole) (Vinberg & Grill-Spector, 2008). This is unlikely as this

influence would only occur for neurons whose receptive fields encompass only the surround RDS and not the image background in their processing. Given the comparatively smaller stimulus size used in our experiment and the limited resolution of fMRI to isolate such neurons, we feel any such effect would have been minimal in our study. Further, a neurophysiology report has also suggested that the decision of which of the planes is the object is already made by V2 (Qiu & von der Heydt, 2005).

In summary, we used fMRI to investigate the processing of depth configurations in the human visual system. We manipulated the disparity of a central target plane as well as the disparity of its surround, and used multivariate classification methods to discriminate between these different configurations. We found that early visual areas and particularly V2 consistently responded to the disparity edge, showing a preference of local disparity processing. We also observed sensitivity to both types of our disparity manipulation in dorsal visual areas, suggesting an ability to recruit multiple sources information for calculations of disparity. Finally, we showed that ventral regions preferred disparity when considered relative to its immediate surround, playing a role in object identification and 3D shape discrimination. Future studies exploring the neural basis of disparity representations should consider not only the configuration of the stimulus that is presented, but also the references that may be used for these disparity calculations.

6. The cortical organisation of binocular disparity

Abstract

Neural responses to similar visual properties are often grouped together and vary systematically along the cortical surface. Although this has been observed for many visual features (e.g., orientation, direction of motion), little is known about the cortical structure of binocular disparity. Here, we sought to identify a topographic map for disparity preferences in the human visual cortex. We presented participants with random-dot stereograms that depicted a depth plane and measured fMRI responses to several magnitudes of disparity. We estimated the preferred disparity of each voxel in two ways. First, we identified the disparity that corresponded to the maximum beta-weight when disparities were presented in a random order. Second, we identified the phase-shift when the depth plane was cycled through the depth field. Although we were unable to find a topographic map of disparity, we consistently observed a large cluster of voxels in intermediate dorsal regions (V3A, V3B/KO, V7) that contained a preference for crossed disparities around our largest magnitude (14 arcmin). When restricted to uncrossed disparities, we revealed sensitivity in this same region for a wider range of disparities (4 – 14 arcmin). These results support the role of the dorsal stream in visually-guided actions, where crossed disparities of this magnitude are useful in guiding movement towards an object.

6.1 Introduction

Binocular disparities are formed from the small differences in the images of each eye and are used to perceive depth in our environment. Individual neurons responsive to disparity have been observed throughout the visual cortex (Gonzalez & Perez, 1998; Neri, 2005; Parker, 2007); however, details of their functional organisation remain elusive. Neurons have been shown to exhibit columnar organisation for several types of visual features, where neurons within one column (perpendicular to the cortical surface) all contain similar preferences for that feature. Moreover, the preferences of these columns are observed to systematically vary for neighbouring columns along the cortical surface and create a topographic map. This has been observed with single-unit recording and optical imaging for a variety of visual features, including ocular dominance (Hubel & Wiesel, 1977), orientation (Blasdel, 1992; Hubel & Wiesel, 1977), spatial frequency (Tootell, Silverman, & De Valois, 1981), direction of motion (Weliky, Bosking, & Fitzpatrick, 1996) and indeed also for binocular disparity (Chen, Lu, & Roe, 2008; DeAngelis & Newsome, 1999).

Clusters of disparity-sensitive neurons were observed to span ~1 mm along the cortical surface in MT interspersed among similar-sized patches of neurons that were not disparity-selective (DeAngelis & Newsome, 1999). The spatial layout of orientation preferences showed a similar distribution, with a 180° change in orientation representing ~1 mm of the cortical surface (Hubel & Wiesel, 1977) and has been recently mapped in V1 using high-field fMRI (Yacoub, Harel, & Ugurbil, 2008). It is therefore conceivable that we could observe a topographic map for disparity using human fMRI. While this has been observed in the thick stripe of V2 and in dorsal region MT, a host of neuroimaging evidence has emphasised the role of other, intermediate dorsal visual areas (Backus et al., 2001; Cottereau et al., 2011, 2012a; Georgieva et al., 2009; Minini et al., 2010; Preston et al., 2008; Tsao et

al., 2003). In particular, V3A is consistently regarded as highly sensitive and is thought to be central to disparity processing, such that it may be better suited to display a topographic map for disparity.

In this study, we varied the depth of a disparity-defined plane and used fMRI to examine the structure of cortical responses in the human visual cortex. In the first experiment, we presented several disparities in a block-design and estimated the preferred disparity of each voxel by extracting the maximum beta-weight from the general linear model (GLM). In a second experiment, we presented disparity planes whose position gradually cycled through the depth field and performed a phase-encoding analysis to reveal the preferred disparity of each voxel. We did not observe a fine spatial organisation of disparity processing in any visual area, but observed a preference towards the larger disparities that were presented in several dorsal visual areas (V3A, V3B/KO, V7) revealing a bias in the representation of disparity at the resolution of the voxel.

6.2 Methods

6.2.1 Participants

Five participants (one female) completed three experimental sessions on separate days. All participants were recruited through the University of Birmingham, including one of the authors. Mean age was 26.6 years (range, 22 – 33 years). Participants had normal or corrected-to-normal vision with no deficits in colour vision and were paid for their participation. Prior to scanning, all participants were screened for stereo deficits in the laboratory to ensure stimuli presented in the study were above observers' detection thresholds. The experiment was granted ethics approval by the University of Birmingham STEM committee and all participants provided written informed consent.

6.2.2 Stimuli

We used stimuli consisting of dense (25 dots/deg^2) correlated random dot stereograms (RDS) presented in a circular aperture (12° diameter). The stereogram was made up of black and white dots, where each dot was given a Gaussian luminance profile with a diameter (at half-height) of 0.15°. Within this region, four sectors of a smaller circular aperture (10° diameter) were given nonzero disparities. Each sector subtended an angle of 75° and was positioned between two reference sectors (angle, 15°) that contained zero disparity and were centred on the horizontal and vertical axes. To minimise the effects of adaptation, the stimulus was rotated from this position up to a maximum of ±2 angular degrees on each trial. For Experiment 1, each target sector was given a disparity of ±2, ±6, ±10, or ±14 arcmin. The two sectors to the left of fixation contained the same disparity and had a different sign but were equal in magnitude to the two sectors on the right of fixation (i.e., disparity left = – disparity right). For Experiment 2, all four target sectors contained the same disparity which ranged from ±2 to ±16 arcmin, in steps of 2 arcmin. The background surrounding the RDS was filled with a grid of black and white squares which could be used as a clear reference and encouraged vergence to the plane of the screen. The fixation marker was a hollow white square (0.55° side length) that was presented in the centre of the stereogram. Vertical and horizontal nonius lines (length 0.25°) were attached to the fixation marker to promote vergence at the plane of the screen. We restricted presentation of dots to outside of a circular region (1.5° diameter) centred at fixation to reduce interference from the stimulus on binocular fusion. Examples of stimuli with corresponding illustrations are shown in **Figure 6.1a** and **Figure 6.1b** for the two experiments, respectively.

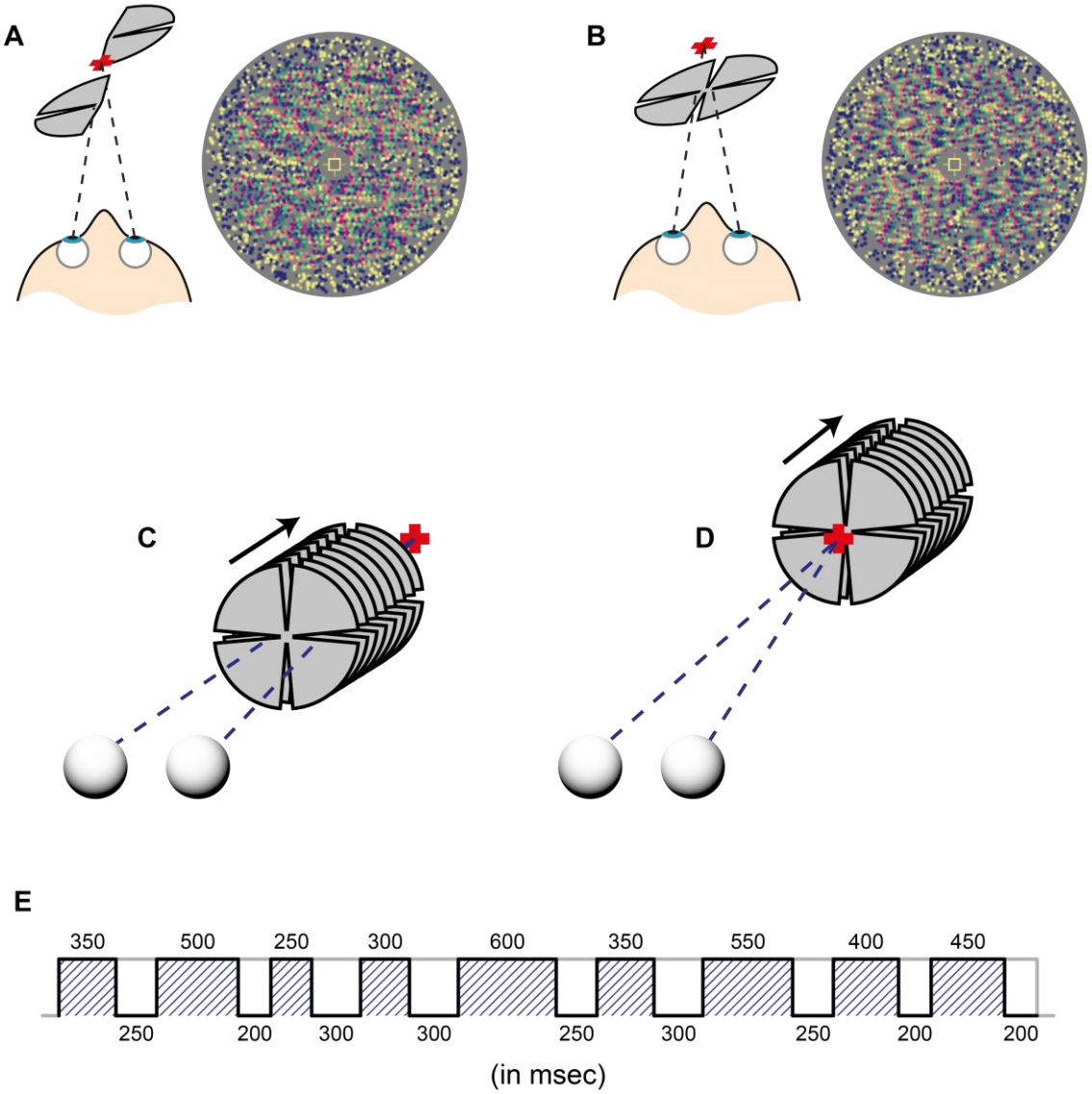


Figure 6.1: Example stimuli from both experiments and the design of Experiment 2. (A & B) An illustration and random-dot stereogram of a stimulus presented in Experiment 1 (A) and Experiment 2 (B). (C & D) The design of two of the run types in Experiment 2. The stimulus was presented in 6 s blocks at each disparity, starting closest to the observer and shifting 2 arcmin in depth away from the observer for each subsequent block. At the end of each cycle, a scrambled block was presented which contained dots from all of these disparities. (E) An example of the permutation of stimulus durations during one stimulus block in Experiment 2. Each block consisted of these stimulus presentation and inter-stimulus interval lengths, but their order was permuted for each block. The grey line represents the length of the block, 6 sec.

6.2.3 Design

Experiment 1. Each run was comprised of eight different stimulus conditions corresponding to the eight different disparity configurations. The blocked presentation of each condition lasted

16 s. Within a single block, 16 different RDS stimuli were presented (500 ms stimulus on; 500 ms stimulus off). Each condition block was repeated three times per experimental run, and conditions were presented in a randomised, counterbalanced order across runs and subjects. Each scan lasted 416 s and included two 16 s fixation periods (one in the beginning and one at the end of the run). Fixation periods consisted of only the fixation square and the background being visible. Participants completed a single session of 9 – 10 scans. During all runs, observers performed a nonius alignment test (McKee & Mitchison, 1988; Popple et al., 1998; Ukwade, Bedell, & Harwerth, 2003). A vertical line (length 0.22°) was presented inside the fixation square in one eye on 25% of trials slightly offset from centre. Participants were asked to report its horizontal position ('left' or 'right') with respect to the nonius line presented to the alternative eye. Prior to the commencement of scanning, participants viewed additional stimuli (30 trials per condition) of the ±2 and ±6 arcmin disparity configurations and were asked to judge whether the sectors on the left or the right side of fixation were perceived as nearer. All participants performed with a high level of accuracy (mean, 99.5%) and could adequately perceive the disparity-defined depth in the stimulus.

Experiment 2. Each run consisted of eight repetitions of a duty cycle that systematically varied depth position. The duty cycle lasted 58 s and consisted of eight disparities presented in blocks of 6 s each and a 'scrambled' block where all dots contained randomly-allocated disparities from the other disparity conditions. This scrambled block lasted 10 s and was used to reduce the circularity of disparities towards the end of the cycle affecting the response of disparities at the start of the next cycle. Each scan lasted 496 s, including a 16 s fixation period that was presented both at the start and end of the run. We performed four different types of runs. For one run type, the disparity of the first block was nearest to the observer (-16 arcmin) and the disparity of each following block decreased in

steps of 2 arcmin towards fixation, to a minimum disparity of -2 arcmin (**Figure 6.1c**), and was followed by a scrambled block. The second run type reversed this procedure, where the scrambled block was presented first, and then stimulus blocks started at -2 arcmin and moved towards the observer using the same disparities. During analysis, these runs were combined after reversing the fMRI data contained in this second run type. This corrected for differences in the phase caused by differences in the delay of the hemodynamic response across visual areas (Kastrup, Kruger, Glover, Neumann-Haefelin, & Moseley, 1999). We then presented a further two runs by shifting all disparities to the equivalent depths behind fixation and repeating this procedure (e.g., **Figure 6.1d**). One session contained two repeats of these four run types, and participants completed two sessions on different days.

To remove regularity in stimulus presentation and minimise aliasing in the higher frequencies of the Fourier spectrum, we varied the duration of stimulus presentations and inter-stimulus intervals (that displayed only the background and fixation) and permuted their order for every block. Each stimulus was presented for between 250 – 600 ms (in increments of 50 ms). Inter-stimulus intervals separated all stimulus presentations and also varied in duration (200, 250, 300 ms). Stimulus blocks were restricted to six seconds by using one of each of the above stimulus durations and an additional presentation of 350 ms, with three repeats of the varying inter-stimulus intervals (see **Figure 6.1e** for an example of the permutation of one block). Scrambled blocks lasted ten seconds and contained two stimulus presentations from each of these durations, except for 550 ms which was only presented once, and five repeats of the varying inter-stimulus interval lengths.

As before, observers completed a nonius alignment test during all runs, with the additional requirement that the vernier stimulus did not appear on consecutive stimulus presentations to allow sufficient time for the participant's response. Immediately prior to

scanning, participants viewed additional stimuli (30 trials per condition) at the ±2 and ±4 arcmin disparity conditions and were asked to judge whether the critical stimulus was in front or behind of fixation. Stimulus duration was randomised as in the experiment. All participants performed with a high level of accuracy (mean, 98.4%) and were able to adequately perceive disparity of these smaller conditions.

6.2.4 fMRI data acquisition

The study was performed in a 3-Tesla Philips Achieva MRI scanner at the Birmingham University Imaging Centre. We used an eight-channel SENSE head coil to collect echo-planar imaging (EPI) and T1-weighted anatomical (1 × 1 × 1 mm) data. EPI data [echo time (TE), 34 ms; repetition time (TR), 2000 ms] were acquired from 28 slices (voxel size, 1.5 × 1.5 × 2 mm) positioned close to the coronal plane. Localisers were obtained in a separate session using the same EPI sequence. We used a pair of video projectors (JVC D-ILA SX21) to display stereoscopic images, with each projector fitted with a unique interference filter (INFITEC, GmBH) that distributed the wavelengths of visible light between the two projectors. The images from each projector were optically combined using a beam-splitter cube and passed through a waveguide into the scanner room. Stimuli were projected onto a translucent plastic screen located behind the head coil and inside the bore of the scanner. Participants viewed the screen via a mirror positioned on the head coil angled at 45°, with an optical path length of 65 cm. Since the INFITEC filters produced negligible overlap between the spectra for each projector, there was minimal crosstalk between the projected images for a participant wearing a pair of corresponding filters. Unique stimuli were pre-generated for each participant using C#, and the experiment was performed using MATLAB (The MathWorks, Natick, MA) and the PsychToolBox 3 extension (Brainard, 1997; Pelli, 1997).

For each participant, we identified regions of interest (ROIs) from independent data in a separate localiser scan. We used a rotating wedge stimulus and expanding concentric rings to define the borders of early retinotopic areas (V1, V2, V3v, V4, V3d, V3A and V7) (Aguirre et al., 1998; DeYoe et al., 1996; Sereno et al., 1995). In particular, V4 was defined as the ventral region adjacent and inferior to V3v containing a full hemifield representation (Tootell & Hadjikhani, 2001; Tyler et al., 2005) and V7 was defined as the dorsal region adjacent and anterior to V3A, also containing a full hemifield representation (Tootell et al., 1998; Tsao et al., 2003; Tyler et al., 2005). We identified higher dorsal regions [V3B/kinetic occipital area (KO), human motion complex (hMT+/V5)] and the ventral lateral occipital region (LO) from additional localiser scans. Area V3B/KO (Dupont et al., 1997; Zeki et al., 2003) was defined anatomically as the dorsal region adjacent and lateral to V3A with which it shared a foveal representation (Tyler et al., 2005), and functionally as the set of voxels that responded significantly more ($p < .001$) to kinetic boundaries than to the transparent motion of black and white dots without clear borders. Area hMT+/V5 was defined as the region in the lateral temporal cortex that responded significantly more ($p < .001$) to an array of coherently moving dots than to an array of static dots (Zeki et al., 1991). The lateral occipital (LO) area was identified as the region in the lateral occipito-temporal cortex that responded significantly more ($p < .001$) to images of objects and shapes than to scrambled versions of these images (Kourtzi & Kanwisher, 2000, 2001). During all localiser scans, participants performed an attentionally demanding task on the fixation point, except for the LO localiser in which they had to respond if the same image was presented consecutively.

6.2.5 fMRI data analysis

We processed MRI data using BrainVoyager QX (BrainInnovation, Maastricht, the Netherlands). For each participant, anatomical scans were transformed into Talairach space and used for 3D cortex reconstruction, inflation, flattening and the segmentation of gray and white matter. Preprocessing of functional data included head movement correction (translation and rotation), slice scan time correction, and removal of low-frequency drifts and linear trends through temporal high-pass filtering (3 cycles). Functional runs were aligned to the participant's anatomical data and then transformed into Talairach space. Consecutive volumes that contained head movement greater than 1 mm of translation or 1° of rotation were excluded from further analysis.

Experiment 1. Responses in left and right hemispheres were analysed separately since the disparity sign differed between stimuli to the left and right of fixation. Where necessary, these results were later combined for corresponding disparities in left and right hemispheres such that we could analyse results per disparity, rather than by the original configuration of the stimulus. We normalised the fMRI time course through a transformation to z-scores and used a two-gamma model to simulate the hemodynamic response function. We computed the GLM and assigned a beta-weight to each disparity condition. All beta-weights were normalised by subtracting the mean weight from each voxel. We estimated the disparity preference of each voxel through finding the disparity condition that corresponded to the largest beta-weight. To ensure reliability of our results, we split the data from each session into odd and even runs, and computed the GLM for each half-session. We presented voxels where the maximum beta-weight of both half-sessions corresponded to the same disparity condition. Specifically, we overlaid these voxels onto a cortical mesh surface and flatmap for each participant with a colour that represented the disparity preference of the voxel.

Experiment 2. The fMRI time series of all duty cycles of the same run type were concatenated for each participant and a discrete Fourier transform (stimulus frequency, 1/58 Hz; sample frequency, 1/2 Hz) was performed on each voxel. This generated phase values that indicated the disparity most likely to correspond with the observed phase shift in that voxel. The phase values of runs which began with a scrambled block and had the disparity move towards the observer were flipped in Fourier space and averaged with runs of the same disparities that moved in the opposite direction. This resulted in two phase values per voxel: one for disparities in front of fixation and one for disparities behind fixation. Equivalently, the powers of all frequencies along the Fourier spectrum were averaged between runs that moved in opposite directions, providing the amplitude, phase, signal, and noise and allowing a calculation of the F-statistic (magnitude of stimulus frequency / mean of the magnitude of the other frequencies). Higher harmonics of the stimulus frequency potentially reflected a response to multiple conditions and were excluded from our calculations. Phase values for significant voxels ($F_{(2, 461)} > 6.96$, $p < .001$) were converted to colours on a spectrum and represented preferences to different disparities. These maps were overlaid onto a cortical mesh surface and displayed on a flatmap for each participant.

6.3 Results

We presented participants with random-dot stereograms that depicted a depth plane while we simultaneously measured fMRI responses across the visual cortex. To investigate the cortical organisation of disparity and whether there was a topography that could be observed along the cortical surface, we estimated the preferred disparity of each voxel in two experiments. In particular, we investigated whether such a pattern existed in early visual areas (V1, V2), ventral regions (V3v, V4, LO) and dorsal regions (V3d, V3A, V3B/KO, V7, hMT+/V5) that

have previously been implicated in the processing of disparity information in the human brain (Backus et al., 2001; Minini et al., 2010; Neri et al., 2004; Rutschmann & Greenlee, 2004). We predicted that such a topography would be observed in intermediate dorsal regions which frequently display highly sensitivity to disparity information (Backus et al., 2001; Cottereau et al., 2011, 2012a; Georgieva et al., 2009; Tsao et al., 2003). Instead, we observed a strong bias in these regions towards a small subset of the disparities presented.

6.3.1 Results of the MVPA classifier in discriminating depth position

First, to ensure our stimulus created detectable changes in neural activity, we used an MVPA classifier to discriminate between different disparity conditions in each ROI (details of this method are described in previous chapters, see: *5.2.5 fMRI data analysis*). In this instance, we used 300 voxels per ROI (or highest number available) and generated training and test patterns for all 8 disparity conditions presented in Experiment 1, thereby performing an 8-way classification. **Figure 6.2** reveals the mean prediction accuracy for each ROI (chance performance, 0.125), showing that depth position was reliably discriminated for all of our

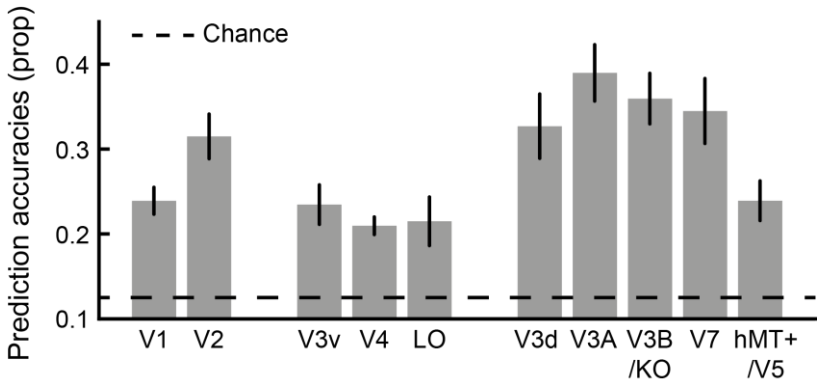


Figure 6.2: An 8-way MVPA classification for disparities presented in Experiment 1. Error bars are standard error. The dotted line indicates chance performance of the classifier (0.125 for an 8-way classifier).

measured visual areas. We therefore conclude that the stimulus and depths we used were appropriate and potentially capable of driving neural changes relating to the depth percept.

6.3.2 Disparity maps from the GLM beta-weights

In the first experiment, we presented depth planes at one of eight disparities using a block-design and measured fMRI responses. We computed the GLM for each participant and recorded the disparity condition that corresponded to the highest beta-weight for voxels throughout the visual cortex. We minimised presentation of noisy data by only displaying voxels that had a consistent preference for a single disparity when the fMRI session was split into odd and even runs. Each stimulus disparity was coded with a unique colour and voxels that showed a reliable preference towards that disparity were presented on the corresponding flatmap for that participant. **Figure 6.3** shows one hemisphere of these maps from each participant. Interestingly, across all participants, we observe a large bias in the representation of the -14 arcmin disparity condition for intermediate dorsal visual areas (V3A, V3B/KO, V7). Most participants also showed a similar bias in later dorsal region hMT+/V5 as well as in ventral region LO. In contrast, there was an increased preference towards crossed disparities closer to fixation (2 and 6 arcmin) for early visual areas, though preferences appeared to be distributed more randomly.

We wanted to examine the structure of disparity throughout the visual cortex, thus we did not use a thresholding procedure to limit activation to specific voxels and merely removed voxels that did not give a consistent signal. It is therefore possible and indeed likely that activity from at least some of these voxels is related to other visual processes (e.g., processing of contours) or due to noise inherent in the fMRI signal. To identify visual areas where signals were more likely to be meaningful, we further investigated the reliability of voxels

within each visual area. In particular, we measured the absolute difference (in arcmin) between the disparity preference for odd runs and the disparity preference for even runs. With the eight disparities tested this gives 8 possible differences, where 0 arcmin indicated the voxel had the same preference for both odd and even runs. The largest possible change in disparity preference was +28 arcmin (-14 arcmin for even runs and +14 arcmin for odd runs)

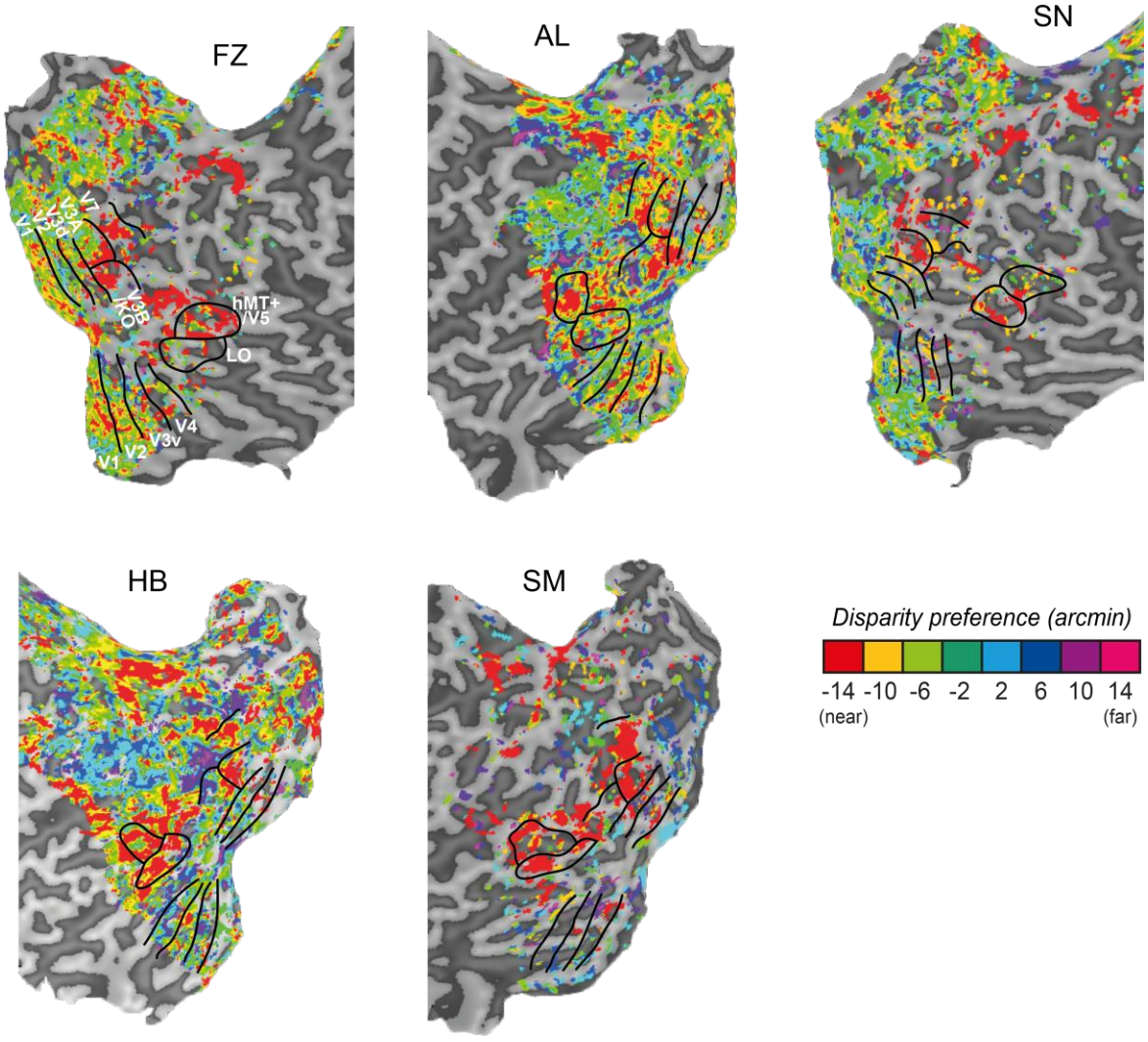


Figure 6.3: Flatmaps of disparity preferences in Experiment 1. Preferences were defined as the disparity condition that corresponded to the maximum beta-weight from the GLM. Each disparity that we presented was given a unique colour so changes in these preferences could be identified across the cortical surface. Only voxels that had the same disparity preference when the fMRI session was divided into odd and even runs were displayed. One hemisphere is presented from each observer, and the labelling of ROIs is shown for participant FZ (top left).

or -28 arcmin (vice versa). We calculated the difference in the disparity preference for all voxels within an ROI and plotted their frequency in **Figure 6.4** as a proportion of the total number of voxels in the ROI. We fit this data with a Gaussian that was restricted to have a mean of zero. This procedure was repeated for all participants and revealed that intermediate dorsal areas (particularly V3A and V7) generally contained the highest proportion of voxels with no difference in disparity preference. In comparison, we observed larger distribution of changes (i.e., a wider Gaussian) in the disparity preference for early visual (V1, V2) and early ventral (V3v, V4) regions. These results suggest the disparity preferences we observed for intermediate dorsal regions are less likely to be due to noise compared to activity in early visual areas. However, it is important to remember this distinction is the proportion of all voxels in the ROI, and does not reveal the reliability of individual voxels. It is therefore possible that a small proportion of voxels in V1 were strongly biased towards a particular disparity, but this signal was hidden amongst a larger proportion of voxels that displayed inconsistent responses.

The analysis of difference in the disparity preferences suggested a higher proportion of voxels in the intermediate dorsal regions gave a consistent response, so we closely inspected the disparity preferences within these regions. We examined the responses in both hemispheres for two participants who displayed the highest reliability in these areas (HB, SM; from **Figure 6.4**). In particular, we displayed the disparity preference for all voxels in these ROIs when all of the runs were analysed together (1 group), or when we restricted voxel presentation to only those where the disparity preference was the same when the data was split into odd and even runs (2 groups), or split into pairs of runs (5 groups) (**Figure 6.5**). We revealed that even with more stringent tests of reliability, responses for some of these voxels continued to emphasise a bias towards the processing of the -14 arcmin disparity condition

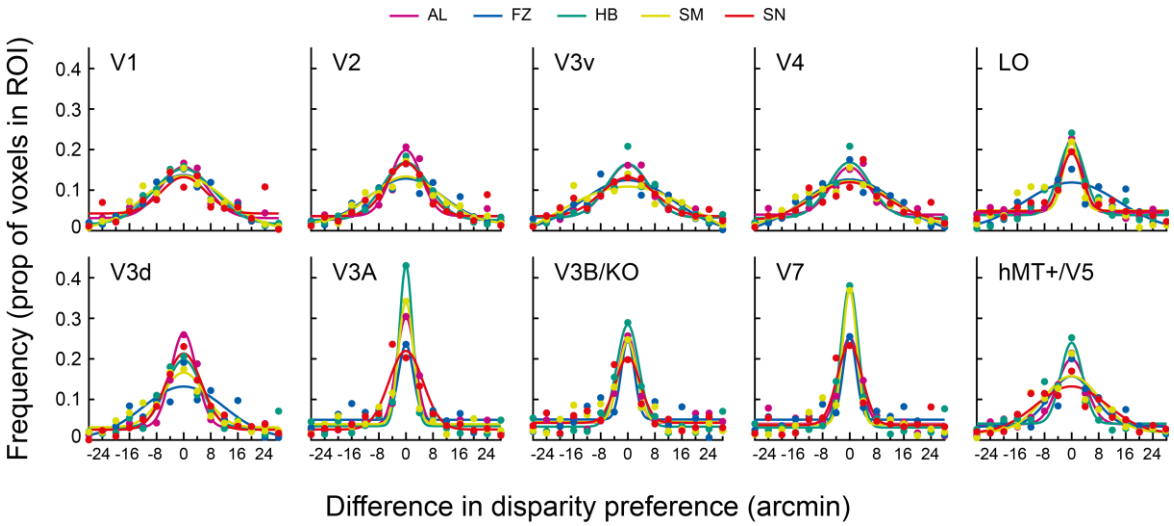


Figure 6.4: Difference in disparity preferences for Experiment 1. This plots the frequency of changes in the disparity preference when obtained for odd runs compared to when obtained for even runs. The frequency is presented as a proportion of voxels within each ROI. Each colour represents the data from a single participant. The data from each participant was fit with a Gaussian that was restricted to have a mean of zero. A positive difference in disparity preference indicates that the preferred disparity was closer to the observer for even runs than it was for odd runs.

(**Figure 6.5**, right column). When we did not restrict the presentation of voxels in these areas, we continued to observe large clusters of voxels that preferred the -14 arcmin disparity condition (**Figure 6.5**, left column). However, we also note a pattern in the disparity preferences of voxels that surround these clusters. Specifically, there was a systematic change in disparity preferences, where voxels that immediately surrounded these clusters showed a single change in disparity preference (i.e., to -10 arcmin) and voxels located next to these similarly changed to the neighbouring presented disparity. Further, we observed an overall bias towards crossed disparities, with only minimal voxels that displayed a preference towards uncrossed disparities. However, the voxels that constitute this pattern were not considered to be reliable (**Figure 6.5**, middle and right columns), and it is possible that this pattern is the result of systematic noise in the fMRI signal.

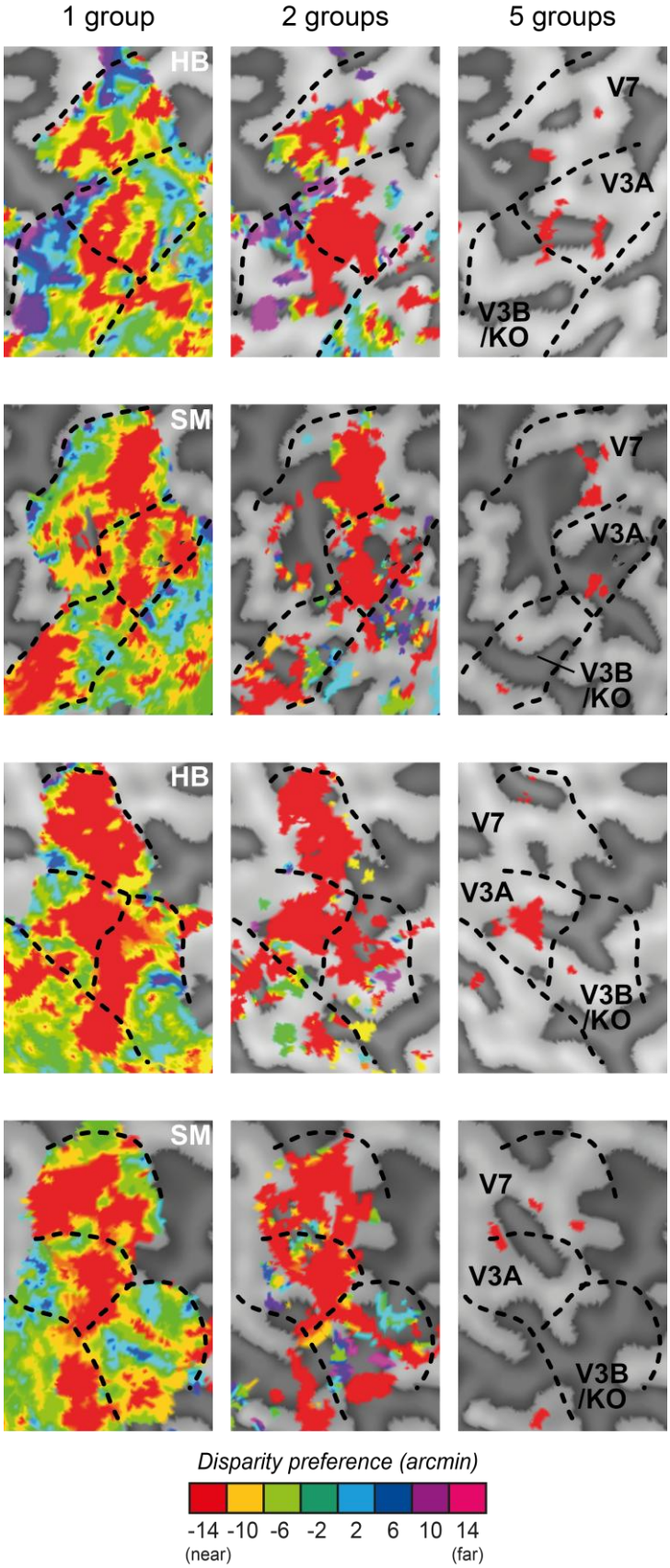


Figure 6.5: Disparity preferences in intermediate dorsal regions in Experiment 1. Each row contains a single flatmap that has been focused on intermediate dorsal regions V3A, V3B/KO and V7. The disparity preference is displayed for all voxels in our ROIs (left column); for voxels that contain the same disparity preference when the fMRI session was split into odd and even runs (middle column), and for when the fMRI session was split into 5 groups (2 runs each; right column). We present the data from both hemispheres of two participants (HB, SM). The left hemispheres of these participants are presented in the top two rows; the right hemispheres are shown in the bottom two rows. Participants are identified in the top right corner of the flatmap presented in the first column. ROIs are labelled in the flatmap presented in the last column.

Thus far, we have identified the preferred disparity for each voxel, but we have not examined how strong this preference is in comparison to the other disparities presented. We therefore examined changes in the beta-weights for the different disparity conditions. **Figure 6.6** shows the mean beta-weights for each ROI (only for voxels which were observed to have the same disparity preference for odd and even runs), and reveals a consistent preference in several regions towards the processing of larger, crossed disparities. As to be expected from the cortical maps of disparity preference (**Figure 6.3**), beta-weights in all dorsal regions and in ventral region LO were greatest for the condition that was nearest to the observer (-14 arcmin). Further, using repeated-measures ANOVA (Bonferroni corrected) we found significant changes in the beta-weights between disparities in early visual area V2, ventral region LO, and in all dorsal regions (V3d, V3A, V3B/KO, V7, hMT+/V5) (**Table 6.1**). The changes to the beta-weights were very similar across these regions: they were generally smallest at 2 or 6 arcmin behind fixation and increased monotonically as the disparity was presented farther away from this depth. In addition, the values of crossed disparities consistently had a beta-weight that was larger than the uncrossed disparity of equal magnitude. This relationship was most evident in intermediate dorsal regions (V3A, V3B/KO, V7), but a similar pattern was observed in the other dorsal visual areas (V3, hMT+/V5) and in ventral region LO. Interestingly, and consistent with the previous analyses, we were unable to identify a clear bias in the disparity preference of early visual (V1, V2) and early ventral (V3v, V4) regions.

These subsequent analyses have all supported our first observation of a bias in the disparity preference of voxels towards larger, crossed disparities. This bias was observed most clearly in a cluster centred in intermediate dorsal regions, and was observed within the boundaries of V3A, V3B/KO and V7 in all of our participants (**Figure 6.3**). Further, we

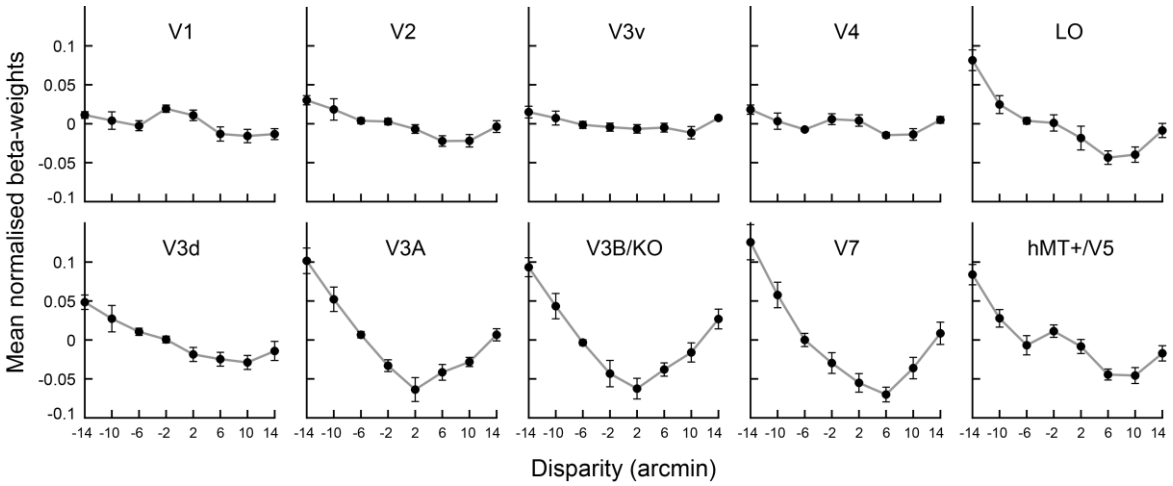


Figure 6.6: Mean normalised beta-weights for each ROI. Each data point represents the mean beta-weights for each ROI and were averaged between subjects. Beta-weights were generated from using all runs in the fMRI session, but only voxels that had the same disparity preference for odd and even runs were included. Error bars are standard error.

observed this same bias towards the -14 arcmin disparity condition for most participants in later dorsal region hMT+/V5, and in later ventral region LO. By examining the mean beta-weights within each visual area (**Figure 6.6**), we confirmed that those for the intermediate dorsal regions showed a strong bias towards near disparities and particularly those of a larger magnitude. These results also suggested a similar bias was present in hMT+/V5 and LO, though this preference for crossed disparities was not as clear as in intermediate dorsal regions. Further it revealed no comparable bias in early visual areas, in line with the appearance of more random pattern of disparity preferences on the cortical surface. By examining the difference in disparity preference when obtaining beta-weights for odd and even runs separately (**Figure 6.4**), we observed a higher proportion of voxels for V3A and V7 that retained the same disparity preference. This suggested that our result in these areas were observed reliably, and was confirmed when we split the fMRI session into 5 groups and still observed a small portion of this cluster that had the same disparity preference in all 5 groups (**Figure 6.5**, right column). The difference in disparity preferences for the other dorsal visual

	F(7,28)	P-value
V1	2.167	.069
V2	4.057	**.003***
V3v	1.405	.243
V4	2.082	.079
LO	9.657	**<.001***
V3d	5.154	**.001***
V3A	16.288	**<.001***
V3B/KO	11.818	**<.001***
V7	14.192	**<.001***
hMT+/V5	11.723	**<.001***

Table 6.1: Repeated-measures ANOVA of beta-weights for different disparities. An asterisk marks a significant difference between different disparities for the beta-weights shown in Fig. 6.5 ($p < .05$, Bonferroni corrected).

areas and in ventral region LO all showed a similar pattern to what was observed in V3A and V7, while early visual (V1, V2) and early ventral regions (V3v, V4) showed broader Gaussian fits, and suggested a distribution more similar to what would be expected from noise.

6.3.3 Disparity maps generated from phase-encoding of depth

In the second experiment, we recorded fMRI responses while presenting a depth plane that systematically moved through the depth field. We used a phase-encoding method to reveal the preferred disparity for each voxel (i.e., the disparity most likely to correspond with the phase-shift described by the fMRI response). We restricted presentation to voxels which were observed to have a significant phase at the frequency of the stimulus. These phase values were converted to colours on a spectrum and displayed on the flatmap for that participant. This procedure was repeated for both stimuli located in front and behind fixation to reveal voxels with a significant disparity preference for either crossed or uncrossed disparities.

Figure 6.7 shows preferences for crossed and uncrossed disparities for one hemisphere in each participant, where a third spectrum is used to show the mean phase of

voxels that had significant disparity preferences for both crossed and uncrossed disparities. Voxels that showed a preference towards the scrambled block (where dots were randomly assigned disparities from the other depth planes in the same run) occurred very minimally, but are shown as medium-grey on respective flatmaps. We observed large variations in the preferred disparities across visual areas and between participants, and were unable to observe any clear topographic structure through differences in these phase values. However, for all participants we observed a large cluster of voxels within intermediate dorsal areas (V3A, V3B/KO, V7) that showed significant phase values for both crossed and uncrossed disparities. Interestingly, this cluster corresponded in an almost-identical location to the cluster we observed for the same participant in the previous experiment. This suggests there is a small region located within intermediate dorsal regions that is particularly sensitive to disparity stimuli and can be observed through a variety of manipulations. The position of this cluster supports previous work that has similarly demonstrated high sensitivity to disparity in the intermediate dorsal pathway (Backus et al., 2001; Cottereau et al., 2011, 2012a; Tsao et al., 2003; Tyler et al., 2006).

We were unable to observe variation in the average phase for this cluster in intermediate dorsal areas, but this may be from combining phases of both crossed and uncrossed disparities in these voxels. We therefore selected two participants (HB, SM) and presented the phases for crossed disparities on a separate map (**Figure 6.8**, top row) to those from uncrossed disparities (**Figure 6.8**, bottom row). We observed some variation in the phase values within this cluster, which appeared largely unstructured and which only covered a subset of the disparities that were presented. Specifically, in both participants we found variations in phase values corresponding to disparities between 10 – 14 arcmin for crossed disparities, and between 4 – 14 arcmin for uncrossed disparities (though there was a bias

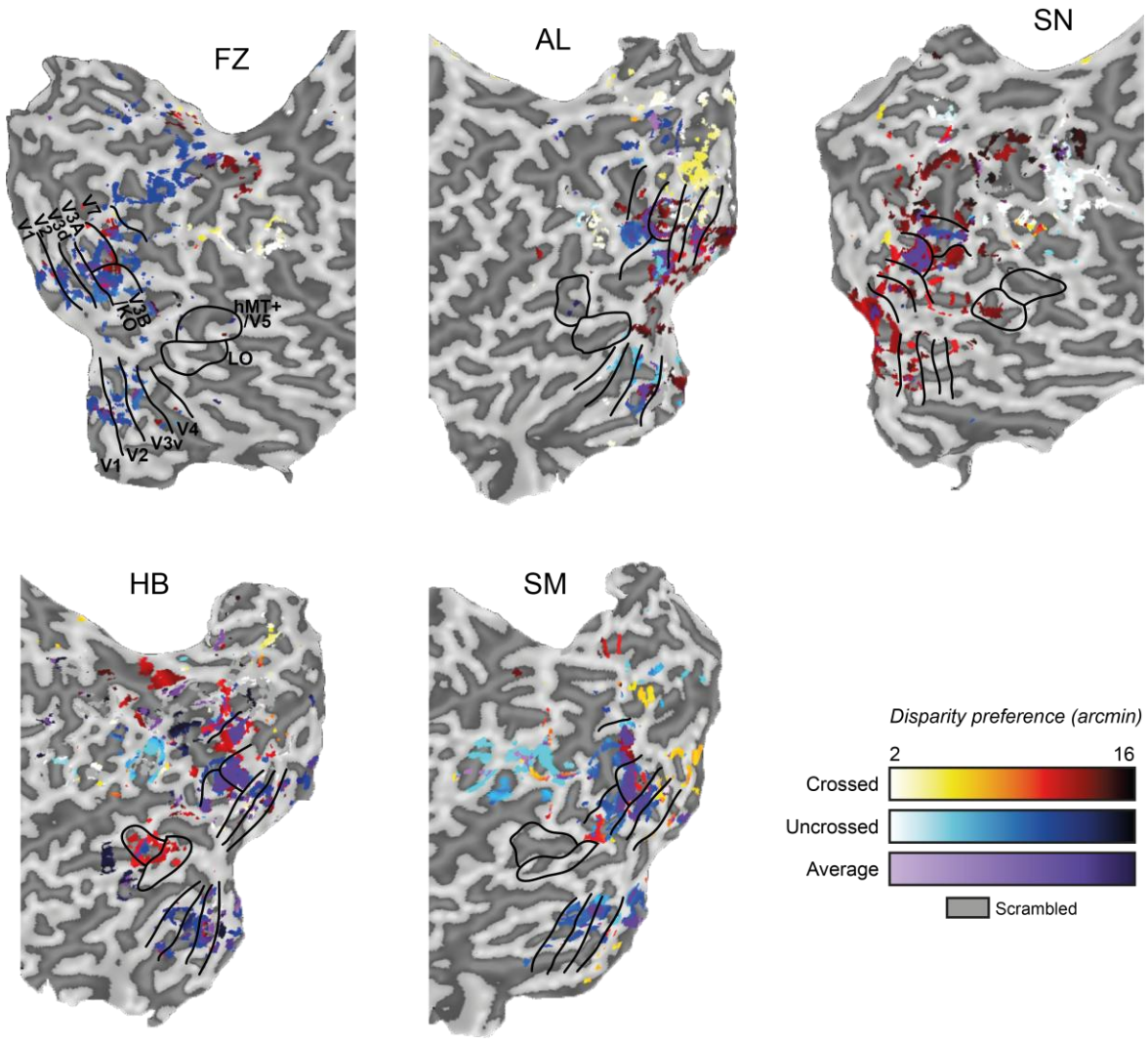


Figure 6.7: Flatmaps of phase values in Experiment 2. One hemisphere of each participant is shown with phase values for runs that contained only crossed disparity (red spectrum) and for runs that contained only uncrossed disparities (blue spectrum). Voxels that contained a significant phase for both uncrossed and crossed disparities are shown in a third, purple spectrum. Voxels that showed a preference towards the scrambled block are shown as medium-grey. The labelling of ROIs is shown for one participant (top left).

towards disparities of greater magnitude). The maps in **Figure 6.8** were representative of the results we observed in all participants and for both hemispheres: they all contained similar unstructured variations in phase values that remained strictly within this range of disparities. However, unlike the previous experiment, responses in hMT+/V5 and LO were observed relatively infrequently and was likely due to the more conservative statistical approach that we used in this experiment.

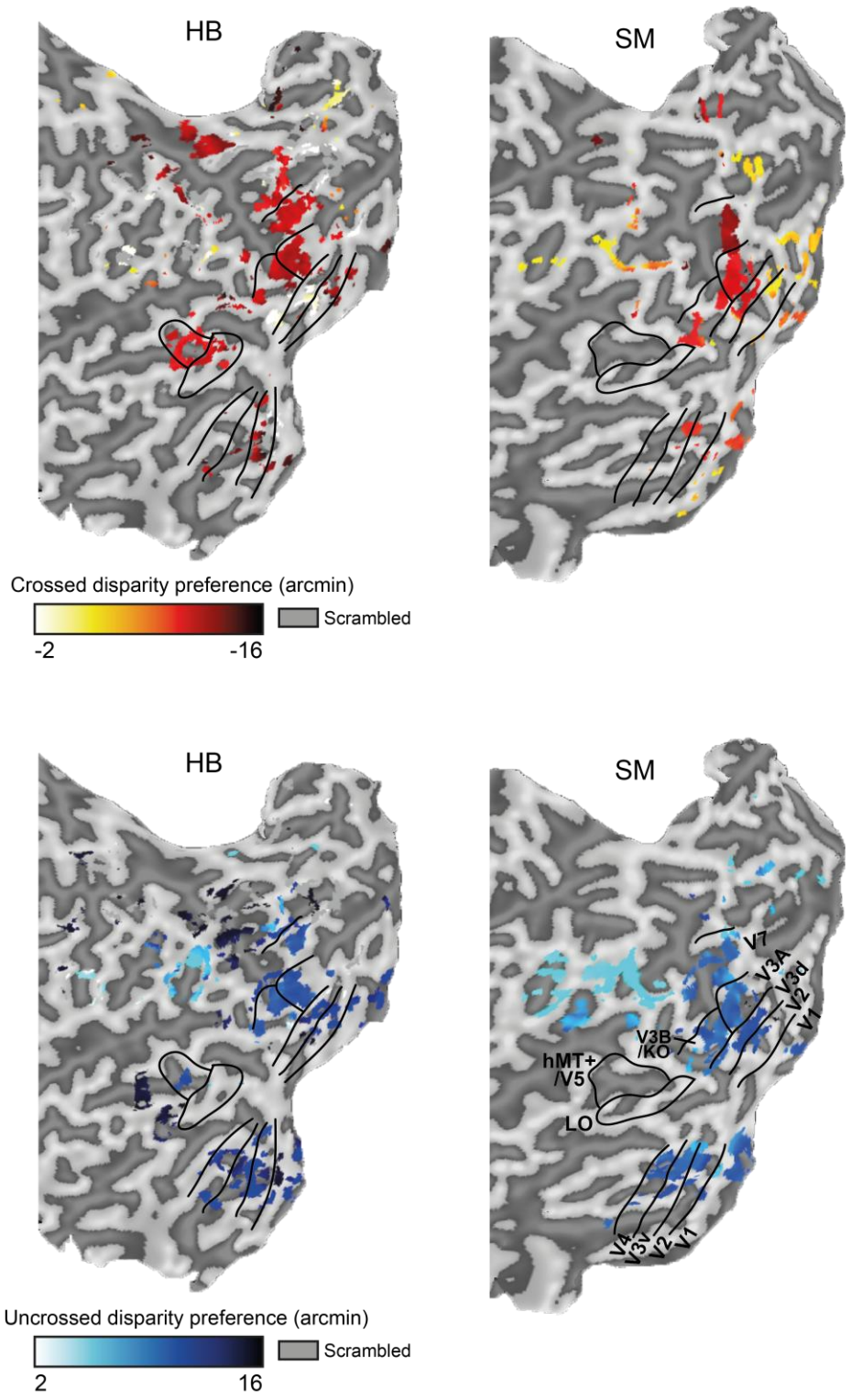


Figure 6.8: Flatmaps of phase values for crossed and uncrossed disparities. Phase values for crossed and uncrossed disparities are presented on separate flatmaps. A single hemisphere from two participants is shown. Flatmaps along the top row show disparity preference to crossed disparities, and flatmaps along the bottom row show the disparity preference to uncrossed disparities. Voxels that showed a preference towards the scrambled block are shown as medium-grey. The labelling of ROIs is shown on the flatmap presented in the bottom right.

6.3.4 Controlling for vergence eye-movements

The confines of the fMRI scanner and our use of spectral filters for stereoscopic presentation meant that we were unable to record changes in eye vergence; however we designed our stimuli to minimise this possibility. In particular, we instructed participants to maintain alignment of vertical and horizontal nonius lines that surrounded the fixation marker throughout all runs, encouraging vergence at the correct depth (the plane of the screen). Further, the stereogram was surrounded by a clear reference pattern that was located at the fixation plane. For Experiment 1, our stimulus planes on one side of fixation contained a crossed disparity while stimulus planes on the other side contained an uncrossed disparity of equal magnitude. This minimised the net stimulus vergence demand.

To ensure participants' vergence state did not become biased away from fixation, participants undertook a vernier task during all scans (Popple et al., 1998). For Experiment 1, vernier results were combined for disparities of the same magnitude, as each disparity magnitude was shown twice: once with the crossed disparity to the left of fixation and once on the right. For Experiment 2, crossed and uncrossed disparities were separated into different runs and were therefore analysed separately. Results of the vernier task indicated that for all conditions in both experiments, there was a minor shift in vergence towards the observer (**Figure 6.9**); however this was very mild in comparison to the changes in disparity, and there was no systematic increase in vergence as the magnitude of the disparity increased. We therefore suggest these were corrective eye-movements related to the stimulus onset and that observers were generally able to maintain stable vergence across conditions despite the vergence demand that accompanies presentation of large disparities. Moreover, as the vergence task at fixation was demanding and presented randomly throughout all conditions, our observed result was unlikely to be caused by changes in the allocation of attention.

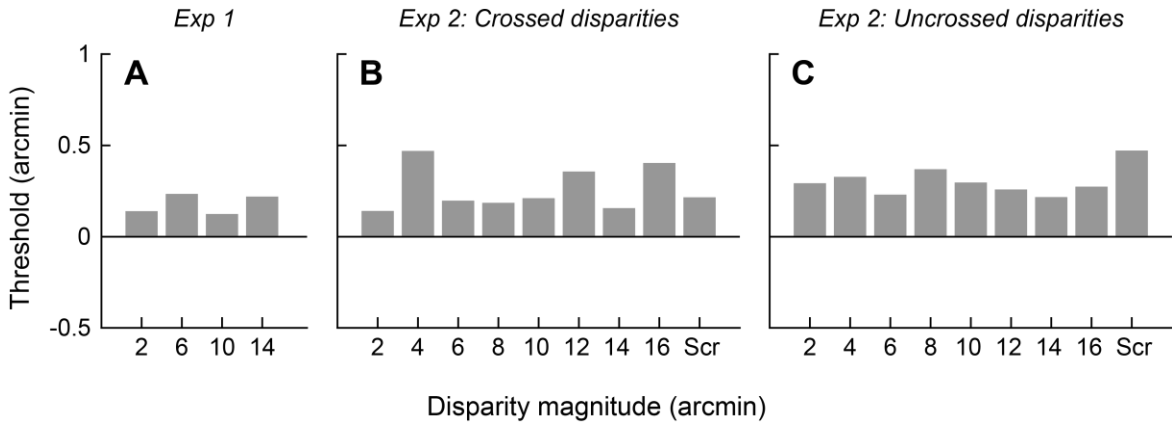


Figure 6.9: Vernier thresholds for both experiments. Results of the vernier task for Experiment 1 (A); runs in Experiment 2 that were comprised of crossed disparities (B) or runs that were comprised of uncrossed disparities (C). Positive values indicate convergence (a shift in vergence towards the observer) and negative values indicates divergence. 'Scr' indicates scrambled blocks consisting of disparity noise stimuli.

6.4 Discussion

We investigated whether we could use fMRI in humans to identify structure in the representation of binocular disparity in the visual cortex. We estimated the disparity preference of voxels in two experiments. First, we identified the disparity that corresponded to the maximum beta-weight when disparities were presented in a random order. Second, we identified the phase-shift when the depth plane was cycled through the depth field. Although we were unable to find a topographic map of disparity, we consistently found a large cluster of voxels spread throughout intermediate dorsal regions (V3A, V3B/KO, V7) that consistently responded to our disparity manipulations. For the first experiment, voxels within this cluster showed a strong preference towards disparities located at -14 arcmin, the condition that was closest to the observer in our experiment. In the second experiment, many voxels co-located to this cluster displayed unstructured variation in their preference of

disparity, though this was still dominated by a preference towards disparities of a greater magnitude.

6.4.1 The range of disparity preferences in intermediate dorsal regions

Through two experiments with independent analyses, we observe an almost-identical section of cortex in the intermediate dorsal pathway of each participant that showed a clear, strong preference to a small range of crossed disparities (10 –14 arcmin). We also observed a wider range in the preferences of uncrossed disparities (4 –14 arcmin) for Experiment 2, and can be explained by the smaller change in the average beta-weights shown in **Figure 6.6**. Here, differences in the beta-weight with changes in disparity are steeper for crossed disparities than for uncrossed disparities, suggesting preferences for crossed disparities would be dominated by a smaller range of disparities.

The preference towards crossed disparities is well-known, and has been observed throughout the visual cortex in the macaque (Adams & Zeki, 2001; DeAngelis & Uka, 2003; Hinkle & Connor, 2001; Prince et al., 2002; Tanabe, Doi, Umeda, & Fujita, 2005; Watanabe, Tanaka, Uka, & Fujita, 2002) and in humans (Cottereau et al., 2011; Preston et al., 2008). The range of preferred disparities is also consistent with values observed in previous work. For example, significant fMRI responses were observed for larger disparities (7.5 – 15 arcmin) when a stimulus consisting of two depth planes was compared to a single depth plane located at fixation (Backus et al., 2001). Similarly, a neuroimaging study revealed that prediction accuracies of an MVPA classifier were highest when discriminating between disparities of larger magnitudes, measuring up to ±15 arcmin (Preston et al., 2008). Using a wider range of disparities (0.5 – 64 arcmin), Cottereau *et al.* (2011) measured EEG responses and observed a peak in the tuning curve between 2 – 16 arcmin for all of their ROIs. As is the case in our

experiments, these results show a bias in the response of underlying neurons that constitute neuroimaging measures, but do not suggest disparities outside of this range are not computed. Indeed, single-unit recording studies have revealed that neurons respond to a much wider range of disparities (Ohzawa, DeAngelis, & Freeman, 1997; Poggio et al., 1988; Prince et al., 2002).

What is the purpose of a bias towards crossed disparities of these magnitudes in the intermediate dorsal pathway? It is established that the dorsal visual pathway plays an important role for orientation and navigation in the visual world (Goodale & Milner, 1992), and it is seen to assist in guiding movements of both the hand (e.g., grasping, reaching) and the eyes (e.g., vergence). We often fixate at a point in the visual field that we are interested in moving towards or interacting with, such as an object we want to reach for. It is therefore sensible that dorsal regions would prioritise crossed disparities in order to aid in comparing the depth between our hand and the point-of-interest, or to identify an obstruction in the path.

However, we acknowledge that our analysis may not have captured the true extent of this bias. In particular, the disparity preference we observed in Experiment 1 corresponded to the largest disparity that was presented. Thus, these regions potentially may have shown a stronger preference towards disparities of larger magnitudes. Examination of a wider range of disparity preferences is necessary to highlight the true peak of this bias. Similarly, it is likely that our observed bias extends into the posterior parietal cortex. Most participants displayed similar disparity preferences along the intraparietal sulcus for Experiment 1 (**Figure 6.3**), and early dorsal regions V3 and V3A have previously been observed to project disparity information to posterior parietal areas (Adams & Zeki, 2001; Nakamura et al., 2001). However, these parietal regions are thought to process higher-order representations of depth (Sakata et al., 2005; Tsutsui et al., 2005). Even though retinotopic organisation has in fact

been observed in parietal regions, these studies used more complex stimulus arrangements (e.g., delayed saccades, visual spatial attention) (Sereno et al., 2001; Silver & Kastner, 2009; Silver, Ress, & Heeger, 2005), making it unlikely that the simple disparity planes we have used would reveal spatial organisation to disparity in these areas.

6.4.2 The structure of binocular disparity in the visual cortex

We were not able to demonstrate a systematic arrangement of disparity preference in any regions of the visual cortex. Of course, it may simply be the case that there is no larger scale organisation of disparity preference. However, topographic maps for disparity selectivity have been observed in the primate visual system for both MT and the thick stripe of V2 (Chen et al., 2008; DeAngelis & Newsome, 1999). Since humans and monkeys have been found to share comparable organisation for other visual features such as ocular dominance (Horton, Dagi, McCrane, & de Monasterio, 1990) and orientation (Obermayer & Blasdel, 1993; Yacoub et al., 2008), it is plausible that this similarity also extends for disparity organisation.

Instead, it is quite likely that our voxel size (1.5 × 1.5 × 2 mm) was insufficient to isolate the responses of columns with specific disparity preferences. For orientation, topographic maps have only been observed with the introduction of high-field fMRI that uses a higher in-plane resolution (0.5 × 0.5 mm^2) (Yacoub et al., 2008). The increased resolution may be necessary for disparity organisation to be observed as the distribution of orientation and disparity preferences have been found to be comparable for the macaque. For example, a 180° change in orientation is distributed over ~1 mm of the cortical surface in V1 (Hubel & Wiesel, 1977). By comparison, the full range of disparity-tuning (varying from far to near and back to far again) was also observed over ~1mm of MT neurons, separated by regions of similar size which did not exhibit disparity selectivity (DeAngelis & Newsome, 1999).

Additionally, neuronal columns that spanned both near and far disparities in V2 were observed to be narrower and have a width of ~0.4mm (Chen et al., 2008). Since organisation in humans is likely to be a factor of 2 – 3 times higher than in the macaque it is not unreasonable to suggest fMRI is close to identifying the spatial clustering of binocular disparity. However, we were unable to observe this organisation for either of our experiments and instead observed a bias towards disparities closer to the observer. Based on the results presented here, we suggest that regions V3A, V3B/KO and V7 would be good candidates to study further and propose that future research samples these regions using high-field fMRI to more likely identify the cortical structure that underlies disparity.

6.4.3 Conclusion

In summary, we sought to identify a topographic map for the representation of disparity on the cortical surface. We presented depth planes to human observers while measuring fMRI activity throughout the visual cortex. We identified the maximum beta-weights in the GLM and performed a phase-encoded analysis to estimate the preferred disparity of each voxel. We did not find systematic variation of disparity preferences in any visual region, but instead observed a cluster of voxels within the boundaries of the intermediate dorsal regions V3A, V3B/KO and V7 that was highly sensitive to several disparity manipulations. Further, this cluster consistently showed a biased representation to a small range of crossed disparities that were located around the largest magnitude that we presented (-14 arcmin). For presentations of uncrossed disparities, disparity preferences were more widespread but were still biased towards larger magnitude disparities. These results are suggestive of a role of the dorsal stream in visually-guided actions, where crossed disparities of this magnitude are useful in guiding movement towards an object.

7. General discussion and conclusions

The work presented in this thesis used human fMRI to gain further insight into the cortical processing of binocular disparity. In particular, our goal was to relate the perception of depth to its corresponding neural signals in the visual cortex. In this chapter, I summarise the main findings from each experimental chapter, I highlight their contributions to the literature and then I examine how the work in this thesis improves our understanding of binocular disparity processing.

7.1 Summary of main findings

7.1.1 Chapter 3: Signal-in-noise and feature difference tasks

Cortical responses to binocular disparity have been observed throughout the visual cortex, with the dorsal and ventral pathways believed to perform distinct computational roles. A series of neurophysiology studies have suggested that specialisation of these pathways can be attributed to differences in disparity magnitude, where the dorsal ventral pathway processes coarse disparities and the ventral stream processes fine disparities. This work used a paradigm involving the discrimination of a target depth plane embedded within a noisy background to measure coarse disparities (i.e., a signal-in-noise discrimination) and the discrimination of small differences in depth to measure fine disparities (i.e., discrimination of feature differences). As both disparity magnitude and perceptual task differs between the two tasks, it was unclear whether differential responses related to the magnitude of the disparity or the task manipulation (internal vs. external noise).

In chapter 3, we decoupled these two properties and investigated cortical areas that related to the perceptual interpretation of depth for these two tasks within a similar range of disparities. Participants completed a near-far depth discrimination task while we simultaneously recorded fMRI measurements. For the signal-in-noise task, we manipulated difficulty by varying the proportion of dots located at the depth of the target plane. For the feature difference task, we manipulated difficulty by varying the relative disparity between the target plane and its surround. We used an MVPA classifier to identify cortical regions whose activity varied in line with perceptual judgments. Specifically, we trained the MVPA classifier to discriminate between fMRI responses for near and far stimuli when the stimulus had the largest signal, and tested its ability to discriminate the target depth across all signal levels, thereby generating 'fMR-metric' functions. We examined the relationship between depth judgments and cortical responses by fitting a scaled version of the mean psychophysical results (restricting the mean and threshold) to the fMR-metric functions in each visual area.

We found we were able to decode clearly-defined depth differences in both ventral (LO) and dorsal (V3A, V3B/KO) cortical areas; however, the relationship between the decoding of these signals and changes in perception was strongest in higher dorsal stream areas. In particular, our results suggest a close relationship between fMRI responses in parietal regions VIPS and POIPS and the perceived depth for both signal-in-noise and feature difference tasks. These results indicate that the distinction between the visual pathways is not caused by the change in perceptual task, and that both tasks require processing in the dorsal visual pathway. We suggest these regions integrate depth information and form the percept of a 3D surface, which is used in later parietal regions to carry out visuomotor tasks such as reaching or grasping.

7.1.2 Chapter 4: Stereo correspondence and binocular perception

Chapter 4 investigated the stage in the visual hierarchy at which the stereo correspondence problem is solved. This task of pairing the same feature in the corresponding images of the two eyes is computationally difficult, as the horizontal offset between the images depends on the distance of the object. This is made even more challenging through random-dot stereograms that contain a multitude of possible matches, but where only one leads to a globally consistent solution. In contrast, reversing the contrast of dots for one of the eyes (i.e., anticorrelated RDS) does not give rise to the perception of depth, as there is no solution which is globally consistent. By examining responses to anticorrelated RDS, it is possible to differentiate visual areas that are responding to local disparity matches to those that respond only when a global solution is perceived (i.e., where the correspondence problem is solved). Using this paradigm, it has been revealed that neurons in V1 respond to local matches (Cumming & Parker, 1997) while neurons at the endpoints of both dorsal and ventral visual pathways (AIP and TEs, respectively) respond to only the global solution that permits stereopsis (Janssen et al., 2003; Theys et al., 2012b).

However, it is unclear where in the visual pathways stereo correspondence takes place. Further, it is unknown how cortical responses vary in line with changes to the proportion of anticorrelated dots. Therefore, in this chapter we recorded fMRI measurements while participants performed a near-far depth discrimination task. We varied the proportion of anticorrelated dots to sample the psychometric function and sought to identify cortical regions that responded in a similar manner. As in Chapter 3, we used an MVPA classifier that was trained to discriminate between fMRI responses for near and far stimuli when the stimulus was perfectly correlated, and tested its ability to discriminate the disparity sign across all proportions of correlated dots, once again generating fMR-metric functions. By scaling the

mean behavioural results of the participants, we examined the extent that classification performance in visual areas matched perceptual responses.

We found that activity in later dorsal regions V7 and VIPS contained activity that closely related to depth judgments, revealing that the stereo correspondence problem is solved at an earlier stage in the dorsal visual pathway than has been observed previously. Furthermore, we revealed a close association between cortical responses and the depth percept at this stage of the visual pathway. This result is consistent with the relative role of other areas in the dorsal pathway, where disparity processing is emphasised in intermediate dorsal regions (e.g., V3A) (Backus et al., 2001; Cottereau et al., 2011, 2012a; Tsao et al., 2003) whereas regions in posterior parietal cortex combine depth estimates from different cues (e.g., Sakata et al., 1998; Taira et al., 2000; Tsutsui et al., 2001). Our finding is therefore compatible with the expected progression of disparity processing in the dorsal visual pathway.

7.1.3 Chapter 5: Depth configurations

Binocular disparities evoke widespread fMRI responses in the visual cortex, indicative of the number of computations that take place between receiving the input from the eyes and binocular depth perception. Investigation of different stimulus configurations reveal the type of processing that is undertaken at that stage of the visual pathway. In this chapter, we investigated the importance of using either fixation or the surround as a reference in disparity calculations.

To do this, we measured fMRI responses and independently manipulated disparity planes presented in a centre-surround configuration. We used a multivariate classifier to discriminate between particular depth configurations and developed three tests to examine the effect of the configuration on visual areas. First, we compared decoding accuracies between

stimulus configurations when the depth of the target plane was held constant relative to the fixation point or the surround. Second, we investigated the effect of increasing the disparity difference between two configurations. Finally, we trained the classifier to discriminate disparity sign using one configuration, and tested the classifier's predictions for depths indicated by a new surround location.

A consistent pattern of responses emerged from these analyses. Early visual areas and particularly V2 contained larger prediction accuracies to a change in the disparity edge than to a change in the disparity of the configuration, consistent with its role in local depth processing (Qiu & von der Heydt, 2005; von der Heydt et al., 2000). Intermediate dorsal areas (V3A, V3B/KO, V7) consistently responded to our manipulations of depth configurations both when depth was considered relative to fixation or the surround, suggesting these regions are able to use both of these as references when calculating disparities. Finally, responses in later ventral region LO obtained higher prediction accuracies when relative to the surround, consistent with its known role in relative disparity processing (Neri et al., 2004; Umeda et al., 2007).

7.1.4 Chapter 6: Cortical organisation of disparity

An organisation of the cortical surface has been observed for a variety of visual features, such as orientation and ocular dominance. Indeed, structure has been observed for binocular disparity in the thick stripes of V2 using optical imaging (Chen et al., 2008) and in MT using single-unit recording (DeAngelis & Newsome, 1999). However, there is no known large-scale topographic map for disparity preferences. In this chapter, we sought to identify cortical structure for binocular disparity in the human visual cortex. We presented depth planes at one of several disparities and measured fMRI responses, estimating the disparity preference of each voxel for two experiments.

In the first experiment, we presented depth planes of different disparities in blocks in a random, counterbalanced order. We extracted the beta-weights from the GLM and identified the disparity condition that contained the largest beta-weight. Voxels were plotted on a flatmap of the cortical surface for that participant, provided they had the same disparity preference when the fMRI session was split into two halves (odd and even runs). This reduced presentation of voxels that were unreliable and likely to be representing noise.

We found a large cluster that lay within the boundaries of intermediate dorsal regions V3A, V3B/KO and V7 which showed a strong preference towards -14 arcmin; the largest crossed disparity that we presented to participants. We closely inspected this finding and observed three properties. First, the disparity preference of this cluster was highly reliable: A portion of this cluster was observed even when the fMRI session was split up into a large number of groups. Second, this bias was more prevalent in intermediate dorsal regions: A higher proportion of voxels in these regions showed a consistent disparity preference when the data was split into odd and even runs than was observed for other visual areas. Third, the bias in these intermediate dorsal regions towards larger, crossed disparities was much stronger than in other visual areas. These analyses combine to suggest a strong, consistent bias in the representation of larger, crossed disparities in the intermediate dorsal visual pathway.

In the second experiment, we presented a depth plane that systematically moved through the depth field. We performed a phase-encoding analysis to indicate the disparity which was most likely to correspond with the observed phase shift in that voxel. Across all observers, we found a cluster in an almost-identical patch of cortex (in V3A, V3B/KO, V7) from the previous experiment that responded significantly to both crossed and uncrossed disparities (which were presented in separate runs). A closer examination of this cluster revealed a preference towards a small range of crossed disparities (10 – 14 arcmin),

replicating our finding in the first experiment. For uncrossed disparities, we observed a larger range of preferences (4 – 14 arcmin), though this cluster was still dominated by preferences towards the larger disparities within this range.

In neither experiment did we observe a topographic map for disparity, but for both experiments we observed a cluster in the intermediate dorsal pathway that showed a clear and consistent bias towards the processing of both larger and crossed disparities. We suggest that the role of this bias is to aid in visually-guided actions, directing an action or movement towards an object located at the point of fixation. Finally, our results suggest that if structural organisation to disparity can be observed throughout the visual cortex, the intermediate dorsal pathway is a good candidate for future research.

7.2 Contributions

This thesis sought out to identify neural responses to binocular depth perception and we have made several advances to the literature. First, by using fine disparities during both signal-in-noise and feature difference tasks we have decoupled the magnitude of the disparity from the perceptual task that was used to investigate it. Our results show that there is no distinction in the visual pathways based on the perceptual task which is performed. This aids in the interpretation of several studies (e.g., Chowdhury & DeAngelis, 2008; DeAngelis et al., 1998; Uka & DeAngelis, 2003, 2006) which have used these tasks to discriminate the cortical networks underlying coarse and fine disparities. Furthermore, by manipulating task performance and measuring fMRI responses concurrently we were able to directly relate changes in the depth percept to changes in cortical activity. This showed that parietal regions use a higher stage of disparity processing that is closely associated with the binocular depth percept. This finding is compatible with previous research that has suggested different depth

cues are combined in parietal regions to form a generalised representation of depth (Sakata et al., 2005; Tsutsui et al., 2005).

Second, using this same approach, we found that activity in later dorsal regions V7 and VIPS contained activity that was closely related to depth judgments. This revealed that the stereo correspondence problem is solved at an earlier stage of the dorsal visual pathway than has been observed previously. This is a significant contribution as it indicates stereo-matching and local filtering is completed by these regions, while areas located higher in the dorsal visual hierarchy perform higher-order functions.

Third, by presenting a range of depth configurations and comparing classification accuracies between these configurations, we used a novel approach to identify differences in the cortical processing of depth. Specifically, we examined changes in prediction accuracies when depth was considered relative to the fixation or the surround and made three contributions. Our finding that V2 processes disparity edges supports previous research (Bredfeldt & Cumming, 2006; Qiu & von der Heydt, 2005; von der Heydt et al., 2000); however the relative contributions of different references were not previously known. Therefore our finding that dorsal visual areas use both fixation and the surround in disparity calculations and that LO preferentially responds when depth is considered relative to the surround, are both novel and interesting contributions. Importantly, these results are also consistent with previous findings in the respective visual pathways: Dorsal visual areas have been shown to be sensitive to a range of disparity stimuli (Backus et al., 2001; Georgieva et al., 2009; Tsao et al., 2003) whereas ventral visual areas compute relative disparities for tasks such as 3D shape processing (Umeda et al., 2007).

Fourth, in our search for a topographic organisation of disparity, we observed a clear and consistent bias within intermediate dorsal regions towards crossed disparities of a larger

magnitude. This finding is a particularly interesting contribution as it indicates that for these regions the representation of disparity is biased at the resolution of the voxel. In revealing this bias, we provide evidence to support the role of visually-guided actions, where crossed disparities of this magnitude are useful in guiding movement towards an object. Further, we reveal that further inspection of these regions using high field fMRI may indeed reveal a topographic map for disparity.

7.3 Responses to binocular disparity throughout the visual cortex

Binocular neurons are first observed in V1 where they perform a local matching task and are apparent throughout the visual cortex, in both ventral and dorsal visual pathways. Recent evidence has suggested that different types of stereo computations are performed between the visual pathways in line with traditional characterisation of the visual streams (Goodale & Milner, 1992). Specifically, disparity in the ventral stream is thought to be used for 3D shape processing and object recognition while disparity in the dorsal stream is thought to relate to navigation and visually-guided actions. Our results are broadly consistent with this proposal.

7.3.1 Responses in early visual areas

In early visual areas, and particularly V2, we observed sensitivity towards disparity edges in line with local processing of disparity signals. Neurons in V1 have been shown to respond to matches that are identical within their receptive field, even if the perceived (global) depth has changed (Cumming & Parker, 2000). Similarly, neurons in V1 have consistently been shown to respond to anticorrelated stereograms, where there is no consistent global solution (Cumming & Parker, 1997). However, responses in some V2 neurons have shown selectivity

to disparity edges (Bredfeldt & Cumming, 2006; Qiu & von der Heydt, 2005; von der Heydt et al., 2000), suggesting this region may begin to represent figure-ground segregation.

7.3.2 Responses in ventral visual areas

For ventral regions, we observed a preferential response to disparity relative to its surround in ventral region LO, suggesting a role in relative disparity processing that is used to discriminate 3D shape. Interestingly, in all of our experiments we observed lower sensitivity to disparity in the ventral stream than in the dorsal stream. One possibility for this is that there is an underlying difference in the way disparity signals are processed in the ventral stream compared to the dorsal stream. Another possibility is that the frontoparallel depth planes that were used throughout this thesis biased activation away from ventral visual areas, as planar stimuli do not contain aspects of 3D shape that the ventral stream is known to favour. However, previous evidence suggests this should not be the case. In one study, the majority of measured V4 neurons have even showed a preference towards frontoparallel planes, where significantly more neurons responded to this stimulus type than either slanted or convex/concave surfaces (Hegde & Van Essen, 2005). Numerous other studies have observed significant activation in our later ventral regions (V4, LO) when using planar stimuli (Bridge & Parker, 2007; Cottereau et al., 2011, 2012a; Neri et al., 2004; Preston et al., 2008), though admittedly these areas have also been observed to respond to 3D shape stimuli (Chandrasekaran et al., 2007; Welchman et al., 2005). Regions further along the ventral stream have shown stronger selectivity towards 3D shape stimuli over equivalent frontoparallel projections in both the inferotemporal cortex (Janssen et al., 2000b) and the ventral premotor cortex (Joly et al., 2009), however we did not record activity from these regions for any of our experiments. We therefore suggest that planar stimuli could potentially

have been observed for our regions of interest across both dorsal and ventral visual pathways, and that our choice of stimulus did not unfairly bias our results away from ventral visual areas.

7.3.3 Responses in dorsal visual areas

By comparison, dorsal regions continuously showed high sensitivity to our disparity stimuli. For example, we observed a cluster of voxels in intermediate dorsal regions that displayed a bias towards the processing of crossed, near disparities which we suggested was used to guide visuomotor actions. Intermediate dorsal regions were also observed to perform disparity calculations with regard to both fixation and surround references, perhaps explaining why these areas are frequently observed to be highly sensitive to disparity signals (Backus et al., 2001; Ban et al., 2012; Cottereau et al., 2011, 2012a; Georgieva et al., 2009; Tsao et al., 2003). We also observed responses in later dorsal regions mirrored perceptual responses and had solved the stereo correspondence problem, though this is not specific to the dorsal stream as: (1) stereo correspondence is observed at the endpoints of both visual pathways (Janssen et al., 2000b; Theys et al., 2012b), and (2) responses relating to the binocular percept has also been found in the ventral stream (Shiozaki et al., 2012; Verhoef et al., 2010).

7.4 Conclusions

In the past half-century since disparity-selective neurons were first observed, large steps have been made in understanding the neural processing of binocular disparity. Non-invasive neuroimaging methods such as fMRI have broadened our capability in studying the cortical networks that underlie binocular disparity signals. With these tools, we have examined several core issues relating to the processing of binocular disparity in the human visual cortex. These

have included investigating the relationship between cortical activity and perception, the significance of a reference plane on depth configurations, and the topographic representation of disparity on the cortical surface.

References

Adams, D. L., & Zeki, S. (2001). Functional organization of macaque V3 for stereoscopic depth. *Journal of Neurophysiology, 86*(5), 2195-2203.

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. *Neuroimage, 8*(4), 360-369.

Anzai, A., Chowdhury, S. A., & Deangelis, G. C. (2011). Coding of Stereoscopic Depth Information in Visual Areas V3 and V3A. *The Journal of Neuroscience, 31*(28), 10270-10282.

Anzai, A., Ohzawa, I., & Freeman, R. D. (1997). Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. *Proc Natl Acad Sci U S A, 94*(10), 5438-5443.

Anzai, A., Ohzawa, I., & Freeman, R. D. (1999). Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *J Neurophysiol, 82*(2), 874-890.

Arthurs, O. J., & Boniface, S. (2002). How well do we understand the neural origins of the fMRI BOLD signal? *Trends in Neurosciences, 25*(1), 27-31.

Backus, B. T., Fleet, D. J., Parker, A. J., & Heeger, D. J. (2001). Human cortical activity correlates with stereoscopic depth perception. *Journal of Neurophysiology, 86*(4), 2054-2068.

Bakin, J. S., Nakayama, K., & Gilbert, C. D. (2000). Visual responses in monkey areas V1 and V2 to three-dimensional surface configurations. *Journal of Neuroscience, 20*(21), 8188-8198.

Ban, H., Preston, T. J., Meeson, A., & Welchman, A. E. (2012). The integration of motion and disparity cues to depth in dorsal visual cortex. *Nature Neuroscience, 15*(4), 636-643.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). Neural mechanism of binocular depth discrimination. *Journal of Physiology-London, 193*(2), 327-&.

Bishop, P. O. (1989). Vertical disparity, egocentric distance and stereoscopic depth constancy: a new interpretation. *Proc R Soc Lond B Biol Sci, 237*(1289), 445-469.

Blakemore, C. (1970). The range and scope of binocular depth discrimination in man. *Journal of Physiology-London, 211*(3), 599-&.

Blakemore, C., & Hague, B. (1972). Evidence for disparity detecting neurons in human visual system. *Journal of Physiology-London, 225*(2), 437-&.

Blakemore, C., & Julesz, B. (1971). Stereoscopic depth aftereffect produced without monocular cues. *Science, 171*(3968), 286-&.

Blasdel, G. G. (1992). Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. *J Neurosci, 12*(8), 3115-3138.

Bradley, D. C., Chang, G. C., & Andersen, R. A. (1998). Encoding of three-dimensional structure-from-motion by primate area MT neurons. *Nature, 392*(6677), 714-717.

Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Res, 36*(9), 1255-1264.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Bredfeldt, C. E., & Cumming, B. G. (2006). A simple account of cyclopean edge responses in macaque V2. *Journal of Neuroscience, 26*(29), 7581-7596.

Bredfeldt, C. E., Read, J. C. A., & Cumming, B. G. (2009). A Quantitative Explanation of Responses to Disparity-Defined Edges in Macaque V2. *Journal of Neurophysiology, 101*(2), 701-713.

Bridge, H., & Parker, A. J. (2007). Topographical representation of binocular depth in the human visual cortex using fMRI. *Journal of Vision, 7*(14).

Britten, K. H., Newsome, W. T., Shadlen, M. N., Celebrini, S., & Movshon, J. A. (1996). A relationship between behavioral choice and the visual responses of neurons in macaque MT. *Vis Neurosci, 13*(1), 87-100.

Britten, K. H., Shadlen, M. N., Newsome, W. T., & Movshon, J. A. (1992). The analysis of visual-motion - a comparison of neuronal and psychophysical performance. *Journal of Neuroscience, 12*(12), 4745-4765.

Buckner, R. L., Goodman, J., Burock, M., Rotte, M., Koutstaal, W., Schacter, D., et al. (1998). Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron, 20*(2), 285-296.

Busettini, C., Miles, F. A., & Krauzlis, R. J. (1996). Short-latency disparity vergence responses and their dependence on a prior saccadic eye movement. *J Neurophysiol, 75*(4), 1392-1410.

Chandrasekaran, C., Canon, V., Dahmen, J. C., Kourtzi, Z., & Welchman, A. E. (2007). Neural correlates of disparity-defined shape discrimination in the human brain. *Journal of Neurophysiology, 97*(2), 1553-1565.

Chen, G., Lu, H. D., & Roe, A. W. (2008). A map for horizontal disparity in monkey V2. *Neuron, 58*(3), 442-450.

Chowdhury, S. A., & DeAngelis, G. C. (2008). Fine Discrimination Training Alters the Causal Contribution of Macaque Area MT to Depth Perception. *Neuron, 60*(2), 367-377.

Cogan, A. I., Lomakin, A. J., & Rossi, A. F. (1993). Depth in anticorrelated stereograms - effects of spatial density and interocular delay. *Vision Research, 33*(14), 1959-1975.

Cormack, L. K., Landers, D. D., & Ramakrishnan, S. (1997). Element density and the efficiency of binocular matching. *J Opt Soc Am A Opt Image Sci Vis, 14*(4), 723-730.

Cottereau, B. R., McKee, S. P., Ales, J. M., & Norcia, A. M. (2011). Disparity-Tuned Population Responses from Human Visual Cortex. *Journal of Neuroscience, 31*(3), 954-965.

Cottereau, B. R., McKee, S. P., Ales, J. M., & Norcia, A. M. (2012a). Disparity-Specific Spatial Interactions: Evidence from EEG Source Imaging. *Journal of Neuroscience, 32*(3), 826-840.

Cottereau, B. R., McKee, S. P., & Norcia, A. M. (2012b). Bridging the gap: global disparity processing in the human visual cortex. *Journal of Neurophysiology, 107*(9), 2421-2429.

Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage, 19*(2), 261-270.

Culham, J. C., Danckert, S. L., DeSouza, J. F. X., Gati, J. S., Menon, R. S., & Goodale, M. A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Experimental Brain Research, 153*(2), 180-189.

Cumming, B. G., & Parker, A. J. (1997). Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature, 389*(6648), 280-283.

Cumming, B. G., & Parker, A. J. (1999). Binocular neurons in V1 of awake monkeys are selective for absolute, not relative, disparity. *Journal of Neuroscience, 19*(13), 5602-5618.

Cumming, B. G., & Parker, A. J. (2000). Local disparity not perceived depth is signaled by binocular neurons in cortical area V1 of the macaque. *Journal of Neuroscience, 20*(12), 4758-4767.

Cumming, B. G., Shapiro, S. E., & Parker, A. J. (1998). Disparity detection in anticorrelated stereograms. *Perception, 27*(11), 1367-1377.

DeAngelis, G. C. (2000). Seeing in three dimensions: the neurophysiology of stereopsis. *Trends in Cognitive Sciences, 4*(3), 80-90.

DeAngelis, G. C., Cumming, B. G., & Newsome, W. T. (1998). Cortical area MT and the perception of stereoscopic depth. *Nature, 394*(6694), 677-680.

DeAngelis, G. C., & Newsome, W. T. (1999). Organization of disparity-selective neurons in macaque area MT. *Journal of Neuroscience, 19*(4), 1398-1415.

DeAngelis, G. C., & Uka, T. (2003). Coding of horizontal disparity and velocity by MT neurons in the alert macaque. *Journal of Neurophysiology, 89*(2), 1094-1111.

DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., et al. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America, 93*(6), 2382-2386.

Dodd, J. V., Krug, K., Cumming, B. G., & Parker, A. J. (2001). Perceptually bistable three-dimensional figures evoke high choice probabilities in cortical area. *Journal of Neuroscience, 21*(13), 4809-4821.

Doi, T., Tanabe, S., & Fujita, I. (2011). Matching and correlation computations in stereoscopic depth perception. *Journal of Vision, 11*(3).

Dupont, P., DeBruyn, B., Vandenberghe, R., Rosier, A. M., Michiels, J., Marchal, G., et al. (1997). The kinetic occipital region in human visual cortex. *Cerebral Cortex, 7*(3), 283-292.

Durand, J. B., Nelissen, K., Joly, O., Wardak, C., Todd, J. T., Norman, J. F., et al. (2007). Anterior regions of monkey parietal cortex process visual 3D shape. *Neuron, 55*(3), 493-505.

Durand, J. B., Peeters, R., Norman, J. F., Todd, J. T., & Orban, G. A. (2009). Parietal regions processing visual 3D shape extracted from disparity. *Neuroimage, 46*(4), 1114-1126.

Erkelens, C. J., & Collewijn, H. (1985). Motion perception during dichoptic viewing of moving random-dot stereograms. *Vision Res, 25*(4), 583-588.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex, 1*(1), 1-47.

Fleet, D. J., Wagner, H., & Heeger, D. J. (1996). Neural encoding of binocular disparity: energy models, position shifts and phase shifts. *Vision Res, 36*(12), 1839-1857.

Freeman, R. D., & Ohzawa, I. (1990). On the neurophysiological organization of binocular vision. *Vision Res, 30*(11), 1661-1676.

Georgieva, S., Peeters, R., Kolster, H., Todd, J. T., & Orban, G. A. (2009). The Processing of Three-Dimensional Shape from Disparity in the Human Brain. *Journal of Neuroscience, 29*(3), 727-742.

Gonzalez, F., & Perez, R. (1998). Neural mechanisms underlying stereoscopic vision. *Progress in Neurobiology, 55*(3), 191-224.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences, 15*(1), 20-25.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience, 27*, 649-677.

Harris, J. M. (2004). Binocular vision: moving closer to reality. *Philos Transact A Math Phys Eng Sci, 362*(1825), 2721-2739.

Haynes, J.-D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7*(7), 523-534.

Heeger, D. J., & Ress, D. (2002). What does fMRI tell us about neuronal activity? *Nature Reviews Neuroscience, 3*(2), 142-151.

Hegde, J., & Van Essen, D. C. (2005). Role of primate visual area V4 in the processing of 3-D shape characteristics defined by disparity. *Journal of Neurophysiology, 94*(4), 2856-2866.

Hinkle, D. A., & Connor, C. E. (2001). Disparity tuning in macaque area V4. *Neuroreport, 12*(2), 365-369.

Horton, J. C., Dagi, L. R., McCrane, E. P., & de Monasterio, F. M. (1990). Arrangement of ocular dominance columns in human visual cortex. *Arch Ophthalmol, 108*(7), 1025-1031.

Howard, I. P., & Rogers, B. J. (2002). *Seeing in depth. Volume 2: Depth perception.* Toronto: I Porteous.

Hubel, D. H., & Wiesel, T. N. (1977). Functional architecture of macaque monkey visual cortex. *Proc R Soc Lond B Biol Sci, 198*(1130), 1-59.

Janssen, P., Vogels, R., Liu, Y., & Orban, G. A. (2001). Macaque inferior temporal neurons are selective for three-dimensional boundaries and surfaces. *J Neurosci, 21*(23), 9419-9429.

Janssen, P., Vogels, R., Liu, Y., & Orban, G. A. (2003). At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. *Neuron, 37*(4), 693-701.

Janssen, P., Vogels, R., & Orban, G. A. (1999). Macaque inferior temporal neurons are selective for disparity-defined three-dimensional shapes. *Proceedings of the National Academy of Sciences of the United States of America, 96*(14), 8217-8222.

Janssen, P., Vogels, R., & Orban, G. A. (2000a). Selectivity for 3D shape that reveals distinct areas within macaque inferior temporal cortex. *Science, 288*(5473), 2054-+.

Janssen, P., Vogels, R., & Orban, G. A. (2000b). Three-dimensional shape coding in inferior temporal cortex. *Neuron, 27*(2), 385-397.

Joly, O., Vanduffel, W., & Orban, G. A. (2009). The monkey ventral premotor cortex processes 3D shape from disparity. *Neuroimage, 47*(1), 262-272.

Julesz, B. (1971). *Foundations of Cyclopean Perception.* Chicago, IL: Chicago University Press.

Kastrup, A., Kruger, G., Glover, G. H., Neumann-Haefelin, T., & Moseley, M. E. (1999). Regional variability of cerebral blood oxygenation response to hypercapnia. *Neuroimage, 10*(6), 675-681.

Kourtzi, Z., & Kanwisher, N. (2000). Cortical regions involved in perceiving object shape. *Journal of Neuroscience, 20*(9), 3310-3318.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293*(5534), 1506-1509.

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.

Krug, K., Cumming, B. G., & Parker, A. J. (2004). Comparing perceptual signals of single V5/MT neurons in two binocular depth tasks. *Journal of Neurophysiology, 92*(3), 1586-1596.

Krug, K., & Parker, A. J. (2011). Neurons in Dorsal Visual Area V5/MT Signal Relative Disparity. *Journal of Neuroscience, 31*(49), 17892-17904.

Kumano, H., Tanabe, S., & Fujita, I. (2008). Spatial frequency integration for binocular correspondence in macaque area V4. *Journal of Neurophysiology, 99*(1), 402-408.

LaConte, S. M. (2011). Decoding fMRI brain states in real-time. *Neuroimage, 56*(2), 440-454.

Logothetis, N. K., & Pfeuffer, J. (2004). On the nature of the BOLD fMRI contrast mechanism. *Magnetic Resonance Imaging, 22*(10), 1517-1531.

Marr, D., & Poggio, T. (1979). Computational theory of human stereo vision. *Proceedings of the Royal Society of London Series B-Biological Sciences, 204*(1156), 301-328.

Maske, R., Yamane, S., & Bishop, P. O. (1984). Binocular simple cells for local stereopsis: comparison of receptive field organizations for the two eyes. *Vision Res, 24*(12), 1921-1929.

Masson, G. S., Busettini, C., & Miles, F. A. (1997). Vergence eye movements in response to binocular disparity without depth perception. *Nature, 389*(6648), 283-286.

Mayhew, J. E., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature, 297*(5865), 376-378.

McKee, S. P. (1983). The spatial requirements for fine stereoacuity. *Vision Research, 23*(2), 191-198.

McKee, S. P., Levi, D. M., & Bowne, S. F. (1990). The imprecision of stereopsis. *Vision Research, 30*(11), 1763-1779.

McKee, S. P., & Mitchison, G. J. (1988). The role of retinal correspondence in stereoscopic matching. *Vision Research, 28*(9), 1001-1012.

Merigan, W. H., & Maunsell, J. H. (1993). How parallel are the primate visual pathways? *Annu Rev Neurosci, 16*, 369-402.

Minini, L., Parker, A. J., & Bridge, H. (2010). Neural Modulation by Binocular Disparity Greatest in Human Dorsal Visual Stream. *Journal of Neurophysiology, 104*(1), 169-178.

Nakamura, H., Kuroda, T., Wakita, M., Kusunoki, M., Kato, A., Mikami, A., et al. (2001). From three-dimensional space vision to prehensile hand movements: The lateral intraparietal area links the area V3A and the anterior intraparietal area in macaques. *Journal of Neuroscience, 21*(20), 8174-8187.

Nakayama, K., & Shimojo, S. (1990). da Vinci stereopsis: depth and subjective occluding contours from unpaired image points. *Vision Res, 30*(11), 1811-1825.

Nassi, J. J., Lyon, D. C., & Callaway, E. M. (2006). The parvocellular LGN provides a robust disynaptic input to the visual motion area MT. *Neuron, 50*(2), 319-327.

Neri, P. (2005). A stereoscopic look at visual cortex. *Journal of Neurophysiology, 93*(4), 1823-1826.

Neri, P., Bridge, H., & Heeger, D. J. (2004). Stereoscopic processing of absolute and relative disparity in human visual cortex. *Journal of Neurophysiology, 92*(3), 1880-1891.

Nguyenkim, J. D., & DeAngelis, G. C. (2003). Disparity-based coding of three-dimensional surface orientation by macaque middle temporal neurons. *Journal of Neuroscience, 23*(18), 7117-7128.

Nikara, T., Bishop, P. O., & Pettigrew, J. D. (1968). Analysis of retinal correspondence by studying receptive fields of binocular single units in cat striate cortex. *Exp Brain Res, 6*(4), 353-372.

Obermayer, K., & Blasdel, G. G. (1993). Geometry of orientation and ocular dominance columns in monkey striate cortex. *J Neurosci, 13*(10), 4114-4129.

Ohzawa, I., Deangelis, G. C., & Freeman, R. D. (1990). Stereoscopic depth discrimination in the visual-cortex - neurons ideally suited as disparity detectors. *Science, 249*(4972), 1037-1041.

Ohzawa, I., DeAngelis, G. C., & Freeman, R. D. (1997). Encoding of binocular disparity by complex cells in the cat's visual cortex. *Journal of Neurophysiology, 77*(6), 2879-2909.

Ohzawa, I., & Freeman, R. D. (1986). The binocular organization of complex cells in the cat's visual cortex. *J Neurophysiol, 56*(1), 243-259.

Orban, G. A. (2011). The Extraction of 3D Shape in the Visual System of Human and Nonhuman Primates. In S. E. Hyman, T. M. Jessell, C. J. Shatz, C. F. Stevens & H. Y. Zoghbi (Eds.), *Annual Review of Neuroscience, Vol 34* (Vol. 34, pp. 361-388). Palo Alto: Annual Reviews.

Orban, G. A., Claeys, K., Nelissen, K., Smans, R., Sunaert, S., Todd, J. T., et al. (2006a). Mapping the parietal cortex of human and non-human primates. *Neuropsychologia, 44*(13), 2647-2667.

Orban, G. A., Janssen, P., & Vogels, R. (2006b). Extracting 3D structure from disparity. *Trends in Neurosciences, 29*(8), 466-473.

Orban, G. A., Sunaert, S., Todd, J. T., Van Hecke, P., & Marchal, G. (1999). Human cortical regions involved in extracting depth from motion. *Neuron, 24*(4), 929-940.

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience, 8*(5), 379-391.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Pettigrew, J. D., Nikara, T., & Bishop, P. O. (1968). Binocular interaction on single units in cat striate cortex: simultaneous stimulation by single moving slit with receptive fields in correspondence. *Exp Brain Res, 6*(4), 391-410.

Pitzalis, S., Sereno, M. I., Committeri, G., Fattori, P., Galati, G., Patria, F., et al. (2010). Human v6: the medial motion area. *Cereb Cortex, 20*(2), 411-424.

Poggio, G. F. (1995). Mechanisms of stereopsis in monkey visual-cortex. *Cerebral Cortex, 5*(3), 193-204.

Poggio, G. F., & Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *J Neurophysiol, 40*(6), 1392-1405.

Poggio, G. F., Gonzalez, F., & Krause, F. (1988). Stereoscopic mechanisms in monkey visual-cortex - binocular correlation and disparity selectivity. *Journal of Neuroscience, 8*(12), 4531-4550.

Popple, A. V., Smallman, H. S., & Findlay, J. M. (1998). The area of spatial integration for initial horizontal disparity vergence. *Vision Research, 38*(2), 319-326.

Preston, T. J., Li, S., Kourtzi, Z., & Welchman, A. E. (2008). Multivoxel Pattern Selectivity for Perceptually Relevant Binocular Disparities in the Human Brain. *Journal of Neuroscience, 28*(44), 11315-11327.

Prince, S. J. D., Cumming, B. G., & Parker, A. J. (2002). Range and mechanism of encoding of horizontal disparity in macaque V1. *Journal of Neurophysiology, 87*(1), 209-221.

Prince, S. J. D., Pointon, A. D., Cumming, B. G., & Parker, A. J. (2000). The precision of single neuron responses in cortical area V1 during stereoscopic depth judgments. *Journal of Neuroscience, 20*(9), 3387-3400.

Qiu, F. T. T., & von der Heydt, R. (2005). Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. *Neuron, 47*(1), 155-166.

Read, J. C. A., & Eagle, R. A. (2000). Reversed stereo depth and motion direction with anti-correlated stimuli. *Vision Research, 40*(24), 3345-3358.

Read, J. C. A., Parker, A. J., & Cumming, B. G. (2002). A simple model accounts for the response of disparity-tuned V1 neurons to anticorrelated images. *Visual Neuroscience, 19*(6), 735-753.

Regan, D., Erkelens, C. J., & Collewijn, H. (1986). Necessary conditions for the perception of motion in depth. *Invest Ophthalmol Vis Sci, 27*(4), 584-597.

Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature, 361*(6409), 253-255.

Rutschmann, R. M., & Greenlee, M. W. (2004). BOLD response in dorsal areas varies with relative disparity level. *Neuroreport, 15*(4), 615-619.

Sakata, H., Taira, M., Kusunoki, M., Murata, A., & Tanaka, Y. (1997). The TINS lecture - The parietal association cortex in depth perception and visual control of hand action. *Trends in Neurosciences, 20*(8), 350-357.

Sakata, H., Taira, M., Kusunoki, M., Murata, A., Tanaka, Y., & Tsutsui, K. (1998). Neural coding of 3D features of objects for hand action in the parietal cortex of the monkey. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 353*(1373), 1363-1373.

Sakata, H., Tsutsui, K. I., & Taira, M. (2005). Toward an understanding of the neural processing for 3D shape perception. *Neuropsychologia, 43*(2), 151-161.

Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic-resonance-imaging. *Science, 268*(5212), 889-893.

Sereno, M. I., Pitzalis, S., & Martinez, A. (2001). Mapping of contralateral space in retinotopic coordinates by a parietal cortical area in humans. *Science, 294*(5545), 1350-1354.

Shadlen, M. N., & Newsome, W. T. (1996). Motion perception: Seeing and deciding. *Proceedings of the National Academy of Sciences of the United States of America, 93*(2), 628-633.

Shikata, E., Tanaka, Y., Nakamura, H., Taira, M., & Sakata, H. (1996). Selectivity of the parietal visual neurones in 3D orientation of surface of stereoscopic stimuli. *Neuroreport, 7*(14), 2389-2394.

Shiozaki, H. M., Tanabe, S., Doi, T., & Fujita, I. (2012). Neural Activity in Cortical Area V4 Underlies Fine Disparity Discrimination. *Journal of Neuroscience, 32*(11), 3830-3841.

Silver, M. A., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends Cogn Sci, 13*(11), 488-495.

Silver, M. A., Ress, D., & Heeger, D. J. (2005). Topographic maps of visual spatial attention in human parietal cortex. *J Neurophysiol, 94*(2), 1358-1371.

Singer, W. (1970). Inhibitory binocular interaction in the lateral geniculate body of the cat. *Brain Res, 18*(1), 165-170.

Snowden, R. J., Treue, S., & Andersen, R. A. (1992). The response of neurons in areas V1 and MT of the alert rhesus monkey to moving random dot patterns. *Exp Brain Res, 88*(2), 389-400.

Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *J Neurosci, 11*(9), 2768-2785.

Srivastava, S., Orban, G. A., De Maziere, P. A., & Janssen, P. (2009). A Distinct Representation of Three-Dimensional Shape in Macaque Anterior Intraparietal Area: Fast, Metric, and Coarse. *Journal of Neuroscience, 29*(34), 10613-10626.

Stevenson, S. B., Cormack, L. K., & Schor, C. M. (1989). Hyperacuity, superresolution and gap resolution in human stereopsis. *Vision Research, 29*(11), 1597-1605.

Stevenson, S. B., Cormack, L. K., Schor, C. M., & Tyler, C. W. (1992). Disparity tuning in mechanisms of human stereopsis. *Vision Research, 32*(9), 1685-1694.

Suzuki, H., & Kato, E. (1966). Binocular interaction at cat's lateral geniculate body. *J Neurophysiol, 29*(5), 909-920.

Swisher, J. D., Halko, M. A., Merabet, L. B., McMains, S. A., & Somers, D. C. (2007). Visual topography of human intraparietal sulcus. *Journal of Neuroscience, 27*(20), 5326-5337.

Taira, M., Tsutsui, K. I., Jiang, M., Yara, K., & Sakata, H. (2000). Parietal neurons represent surface orientation from the gradient of binocular disparity. *Journal of Neurophysiology, 83*(5), 3140-3146.

Takemura, A., Inoue, Y., Kawano, K., Quaia, C., & Miles, F. A. (2001). Single-unit activity in cortical area MST associated with disparity-vergence eye movements: Evidence for population coding. *Journal of Neurophysiology, 85*(5), 2245-2266.

Tanabe, S., & Cumming, B. G. (2008). Mechanisms Underlying the Transformation of Disparity Signals from V1 to V2 in the Macaque. *Journal of Neuroscience, 28*(44), 11304-11314.

Tanabe, S., Doi, T., Umeda, K., & Fujita, I. (2005). Disparity-tuning characteristics of neuronal responses to dynamic random-dot stereograms in macaque visual area V4. *Journal of Neurophysiology, 94*(4), 2683-2699.

Tanabe, S., Umeda, K., & Fujita, I. (2004). Rejection of false matches for binocular correspondence in macaque visual cortical area V4. *Journal of Neuroscience, 24*(37), 8170-8180.

Theys, T., Pani, P., van Loon, J., Goffin, J., & Janssen, P. (2012a). Selectivity for Three-Dimensional Shape and Grasping-Related Activity in the Macaque Ventral Premotor Cortex. *Journal of Neuroscience, 32*(35), 12038-12050.

Theys, T., Srivastava, S., van Loon, J., Goffin, J., & Janssen, P. (2012b). Selectivity for three-dimensional contours and surfaces in the anterior intraparietal area. *Journal of Neurophysiology, 107*(3), 995-1008.

Thomas, O. M., Cumming, B. G., & Parker, A. J. (2002). A specialization for relative disparity in V2. *Nature Neuroscience, 5*(5), 472-478.

Tootell, R. B. H., & Hadjikhani, N. (2001). Where is 'dorsal V4' in human visual cortex? Retinotopic, topographic and functional evidence. *Cerebral Cortex, 11*(4), 298-311.

Tootell, R. B. H., Hadjikhani, N., Hall, E. K., Marrett, S., Vanduffel, W., Vaughan, J. T., et al. (1998). The retinotopy of visual spatial attention. *Neuron, 21*(6), 1409-1422.

Tootell, R. B. H., Silverman, M. S., & De Valois, R. L. (1981). Spatial frequency columns in primary visual cortex. *Science, 214*(4522), 813-815.

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S., & Imbert, M. (1992). Modulation of neural stereoscopic processing in primate area V1 by the viewing distance. *Science, 257*(5074), 1279-1281.

Trotter, Y., Celebrini, S., Stricanne, B., Thorpe, S., & Imbert, M. (1996). Neural processing of stereopsis as a function of viewing distance in primate visual cortical area V1. *J Neurophysiol, 76*(5), 2872-2885.

Tsao, D. Y., Vandluffel, W., Sasaki, Y., Fize, D., Knutsen, T. A., Mandeville, J. B., et al. (2003). Stereopsis activates V3A and caudal intraparietal areas in macaques and humans. *Neuron, 39*(3), 555-568.

Tsutsui, K. I., Jiang, M., Yara, K., Sakata, H., & Taira, M. (2001). Integration of perspective and disparity cues in surface-orientation-selective neurons of area CIP. *Journal of Neurophysiology, 86*(6), 2856-2867.

Tsutsui, K. I., Sakata, H., Naganuma, T., & Taira, M. (2002). Neural correlates for perception of 3D surface orientation from texture gradient. *Science, 298*(5592), 409-412.

Tsutsui, K. I., Taira, M., & Sakata, H. (2005). Neural mechanisms of three-dimensional vision. *Neuroscience Research, 51*(3), 221-229.

Tyler, C. W. (1990). A stereoscopic view of visual processing streams. *Vision Research, 30*(11), 1877-1895.

Tyler, C. W., Likova, L. T., Chen, C. C., Kontsevich, L. L., Schira, M. M., & Wade, A. R. (2005). Extended concepts of occipital retinotopy. *Current Medical Imaging Reviews, 1*(3), 319-329.

Tyler, C. W., Likova, L. T., Kontsevich, L. L., & Wade, A. R. (2006). The specificity of cortical region KO to depth structure. *Neuroimage, 30*(1), 228-238.

Uka, T., & DeAngelis, G. C. (2003). Contribution of middle temporal area to coarse depth discrimination: Comparison of neuronal and psychophysical sensitivity. *Journal of Neuroscience, 23*(8), 3515-3530.

Uka, T., & DeAngelis, G. C. (2004). Contribution of area MT to stereoscopic depth perception: Choice-related response modulations reflect task strategy. *Neuron, 42*(2), 297-310.

Uka, T., & DeAngelis, G. C. (2006). Linking neural representation to function in stereoscopic depth perception: Roles of the middle temporal area in coarse versus fine disparity discrimination. *Journal of Neuroscience, 26*(25), 6791-6802.

Uka, T., Tanabe, S., Watanabe, M., & Fujita, I. (2005). Neural correlates of fine depth discrimination in monkey inferior temporal cortex. *Journal of Neuroscience, 25*(46), 10796-10802.

Ukwade, M. T., Bedell, H. E., & Harwerth, R. S. (2003). Stereopsis is perturbed by vergence error. *Vision Research, 43*(2), 181-193.

Umeda, K., Tanabe, S., & Fujita, I. (2007). Representation of stereoscopic depth based on relative disparity in macaque area V4. *Journal of Neurophysiology, 98*(1), 241-252.

Vanduffel, W., Fize, D., Peuskens, H., Denys, K., Sunaert, S., Todd, J. T., et al. (2002). Extracting 3D from motion: Differences in human and monkey intraparietal cortex. *Science, 298*(5592), 413-415.

Verhoef, B.-E., Vogels, R., & Janssen, P. (2010). Contribution of Inferior Temporal and Posterior Parietal Activity to Three-Dimensional Shape Perception. *Current Biology, 20*(10), 909-913.

Vinberg, J., & Grill-Spector, K. (2008). Representation of shapes, edges, and surfaces across multiple cues in the human visual cortex. *Journal of Neurophysiology, 99*(3), 1380-1393.

von der Heydt, R., Zhou, H., & Friedman, H. S. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Research, 40*(15), 1955-1967.

Watanabe, M., Tanaka, H., Uka, T., & Fujita, I. (2002). Disparity-selective neurons in area V4 of macaque monkeys. *J Neurophysiol, 87*(4), 1960-1973.

Welchman, A. E., Deubelius, A., Conrad, V., Bulthoff, H. H., & Kourtzi, Z. (2005). 3D shape perception from combined depth cues in human visual cortex. *Nature Neuroscience, 8*(6), 820-827.

Weliky, M., Bosking, W. H., & Fitzpatrick, D. (1996). A systematic map of direction preference in primary visual cortex. *Nature, 379*(6567), 725-728.

Westheimer, G. (1979). Cooperative neural processes involved in stereoscopic acuity. *Exp Brain Res, 36*(3), 585-597.

Westheimer, G., & McKee, S. P. (1977). Spatial configurations for visual hyperacuity. *Vision Research, 17*(8), 941-947.

Westheimer, G., & McKee, S. P. (1978). Steroscopic acuity for moving retinal images. *J Opt Soc Am, 68*(4), 450-455.

Wheatstone, C. (1838). Contributions to the physiology of vision: part the first. On some remarkable, and hitherto unobserved, phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London, 128*, 371-394.

Williams, Z. M., Elfar, J. C., Eskandar, E. N., Toth, L. J., & Assad, J. A. (2003). Parietal activity and the perceived direction of ambiguous apparent motion. *Nature Neuroscience, 6*(6), 616-623.

Xue, J. T., Ramoa, A. S., Carney, T., & Freeman, R. D. (1987). Binocular interaction in the dorsal lateral geniculate nucleus of the cat. *Exp Brain Res, 68*(2), 305-310.

Yacoub, E., Harel, N., & Ugurbil, K. (2008). High-field fMRI unveils orientation columns in humans. *Proc Natl Acad Sci U S A, 105*(30), 10607-10612.

Zeki, S., Perry, R. J., & Bartels, A. (2003). The processing of kinetic contours in the brain. *Cerebral Cortex, 13*(2), 189-202.

Zeki, S., Watson, J. D. G., Lueck, C. J., Friston, K. J., Kennard, C., & Frackowiak, R. S. J. (1991). A direct demonstration of functional specialization in human visual-cortex. *Journal of Neuroscience, 11*(3), 641-649.

Zhang, J. X., Meeson, A., Welchman, A. E., & Kourtzi, Z. (2010). Learning Alters the Tuning of Functional Magnetic Resonance Imaging Patterns for Visual Forms. *Journal of Neuroscience, 30*(42), 14127-14133.