THE UNIVERSITY *of* EDINBURGH

The Perception and Processing of Self-Motion Cues

Michael Smith

Doctor of Philosophy
School of Informatics
University of Edinburgh
2013

Abstract

The capacity of animals to navigate through familiar or novel environments depends crucially on the integration of a disparate set of self motion cues. The study begins with one of the most simple, planar visual motion, and investigates the cortical organisation of motion sensitive areas. It finds evidence of columnar organisation in hMT+ and a large scale map in V1. Chapter 3 extends this by using stimuli designed to emulate visual and auditory forward motion. It finds that participants are able to determine their direction with a precision close to that predicted by Bayesian integration. Predictions were made regarding neural processing through a modified divisive normalisation model, which was also used to fit the behavioural adaptation results. The integration of different modalities requires visual and auditory streams to combine at some stage within the sensory processing hierarchy. Previous research suggests the ventral intraparietal region (VIP) may be the seat of such integration. Chapter 4 tests whether VIP does combine these cues and whether the correlation between VIP and the unimodal regions changes depending on the coherence of unimodal stimuli. The presence of such modulation is predicted by some models, such as the divisive normalisation model. The processing of such egocentric self motion cues leads to the updating of allocentric representations, these are believed to be encoded by head direction cells and place cells. The experiment in chapter 5 uses a virtual reality stimulus during fMRI scanning to give participants the sense of moving and navigating. Their location in the virtual environment was decoded above chance from voxels in the hippocampus. No head direction signal was classified above chance from any of the three cortical regions investigated. We tentatively conclude that head direction is considerably more difficult to classify from the BOLD signal, possibly due to the homogeneous organisation of head direction cells.

Acknowledgements

Thanks first to my supervisors, Thomas Wolbers and Mark van Rossum for guiding me through the PhD. My colleagues Mat and Sarah for their company and help and to Katrin Julia for performing flawlessly much of the psychophysics for chapter 3. To the dozens of participants who sat through the hours of dot-motion stimuli, experiments wouldn't happen without them. Frances Crabbe for the expert operation of the fMRI scanner at the CCNi. Magdalena (Leni) Wutte for the fMRI planar motion data in chapter 2 and who made many suggestions for the paper it is based on. The advice and data from countless colleagues and academics, including Alexa Morcom, Emma Wood, David Bett, Paul Dudchenko and Patrick Spooner from the lab. Pascal Belin for suggestions about the auditory cues and Lars Muckli for providing access to the MRI. Thanks also to Peter Kaskan for permission to use the optical imaging results and finally thanks to Lyndsey, my family and my friends for putting up with the PhD and for their support over the years.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Michael Smith*)

Table of Contents

Chapter 1

Introduction

Our ability to move and navigate through familiar and novel environments is an ability we take for granted. Without a second thought, we cycle to and from work, take detours and learn new routes. These abilities often only become apparent by their absence - either through head injury or disease. Many people are unable to travel even to familiar places due to their fear of becoming lost (Sohlberg et al., 2005) and over one quarter of people suffering mild-to-moderate dementia reported getting lost in their own homes (Ballard et al., 1991). Intriguingly the navigation deficits reported in those suffering from Alzheimer's Disease may not be purely memory or spatial-map related, but could be due to difficulties interpreting basic visual features such as optic flow (Tetewsky & Duffy, 1999). It seems likely then that understanding our capacity for navigation will require study of the whole pathway, from basic perception through to integrating different stimuli and finally tracking one's orientation and location.

Figure 1.1: Pink footed geese migrating (photo courtesy gidzy, flickr)

Incredible feats of navigation are found across the animal kingdom; the pink-footed goose's migration, wintering in the UK and returning to Iceland and Greenland during the summer can be easily seen here in Scotland (figure 1.1). Similarly, Atlantic Salmon return to the tributary in which they hatched. Leaping salmon is a common sight in nearby rivers and streams during the autumn months. These

examples of navigation and the more mundane ones of walking without colliding into walls are only possible through the collection of information regarding one's own motion from several, complementary, sensory cues. Examples include optic flow (Gibson, 1950; Britten, 2008), auditory vection (Riecke et al., 2009), vestibular responses and proprioceptive feedback. These cues are, by necessity, combined (e.g. DeAngelis & Angelaki, 2012), to give a unified estimate of the environment.

Navigation, the process of finding and traversing a course to a goal location, can be divided into two strategies; allocentric and egocentric navigation (Hölscher et al., 2004). The allocentric frame of reference is that orientated with respect to the external environment. This is usually considered to be dependent on some form of cognitive map in which the positions and directions of global landmarks are stored. The processing and creation of such maps seems to depend to an extent on the hippocampus and are relatively long-term. Egocentric location and navigation, in contrast, is relative to parts of the body, and are driven by perception, imagery and attention (Byrne et al., 2007). An easy way to see this distinction is through the types of instructions one might follow. For example 'turn right at the pond then left at the house' would not require a map of the environment, and would purely depend on an egocentric process (the recognition of landmarks and movements in an egocentric reference frame). The allocentric process to achieve the same goal would require the relative locations of the goal and landmarks to be provided (for instance on a map) and processed to find a path to the destination. The importance and relevance of these two methods in human navigation have been strongly debated (e.g. Wang & Spelke, 2002; Byrne et al., 2007).

The study in chapter 2 investigates a single component (optic flow) which makes up part of the perception of motion used by egocentric navigation. Similarly, chapters 3 and 4 use simple stimuli and investigate purely egocentric navigation. Allocentric landmarks (i.e. long-term features to allow allocentric navigation) are made deliberately unavailable in the stimuli used. The study in chapter 5 moves into the field of allocentric navigation by investigating head direction and place-activity in a global reference frame and using an environment with stable allocentric landmark cues (both distal and local).

Our ability to use either egocentric or allocentric navigation depends on the perception of cues about self-motion and nearby landmarks. The most obvious sense we use to control self-motion is the vestibular (balance) sensation provided by the otolith organs of the inner ear. However, this cue is combined with other stimuli, including auditory and visual motion cues (such as optic flow). For example, the presence of optic-flow is also used in maintaining balance, shown in a study in which invalid optic flow cues caused young children to fall over (Lee & Aronson, 1974). Given that many stimuli (visual, auditory, vestibular and proprioceptive, e.g. Nardini et al., 2008) are used in self-motion processing, the obvious question is how are these cues used? Is only one used in a give situation, or are they all combined and used simultaneously?

It has been widely found that, for a variety of environmental features, the stimuli are weighted so that the more reliably one is given more weight. The ventriloquist effect offers a clear example of this in action. This effect is regularly experienced in the cinema or when watching television; the visual estimate of the source of the sound (mouth of an actor) overrides the auditory estimate of the sound

source (the location of the speakers). The more certainty a source provides about a feature the more weight that source of information is given.

In humans, psychologists have for many years investigated the behavioural and perceptual aspects of self motion cues. From optic flow (Lee & Aronson, 1974) to auditory vection (Larsson et al., 2004) and the influence of vestibular (DeAngelis & Angelaki, 2012) cues and proprioceptive feedback. Within these sensory areas, psychometric curves (e.g. Battaglia et al., 2003) and anisotropies (Gros et al., 1998) have been described and refined. Much of this work allows indirected suppositions or hypotheses about underlying neural circuits to be made. For example the waterfall after-effect, a well-known optical illusion, was investigated by a PhD student, A. Wohlgemuth, over 100 years ago. In his thesis, (Wohlgemuth, 1911) he suggests that the illusion is caused by the 'fatigue' in one of a pair of neurons, a hypothesis tested fifty years later by (Barlow & Hill, 1963), in which they find that retinal ganglion cells are 'fatigued' over similar time-scales as the perceived after-effect (although the motion-aftereffect is now known to be due to adaptation within the cortex, not the retina). The organisation of paired, opposing preferences may be quite a common feature across cortical regions. Indeed, the results of chapter 2 are almost certainly dependent on the pairing of cortical columns with those of an opposing direction preference. How ubiquitous such organisation is remains to be seen. Being able to infer the organisation of cortical regions using fMRI may allow future studies to explain differences in perception (for example between disease-groups and controls). Chapter 2 also finds direct evidence for motion-preference anisotropies: patterns of preference across the primary visual cortex. The planar-motion stimulus in chapter 2 is only part of the optic-flow cue, and as such represents only a fragment of the stimuli associated with self-motion, but is a good place to begin the study of self-motion cues.

Over the last fifty years, studies such as Barlow & Hill (1963) have investigated the neurological underpinnings of such experiences by using electrophysiology to probe the responses of neurons *in vivo* whilst presenting stimuli to the animal. For example, the study by Schlack et al. (2005) found overlapping receptive fields for auditory and visual stimuli in the Ventral Intraparietal Region (VIP), a region believed to be key to integrating cues, including visual and auditory stimuli. Complementing the electrophysiological experiments are psychophysics studies investigating the perception of stimuli through behavioural tasks. These suggest that some form of 'optimal' integration takes place in our integration of various cues. The importance of correct integration can be seen, for example, in its involvement with older people suffering falls (Hay et al., 1996). The hypothesis of optimal integration is tested for the integration of visual and auditory self-motion cues, in chapter 3.

Since the early 90s functional Magnetic Resonance Imaging (fMRI) has been answering questions about cortical processing which have previously been in the domain of electrophysiology. Standard (General Linear Model, GLM) fMRI has allowed the localisation of neural activity in humans to be associated with different stimuli or cognitive tasks, while more advanced methods have allowed details of the processing taking place to be investigated. For example multivoxel pattern classification allows patterns within the brain activity to be compared and 'decoded', whilst methods such as structural equation modelling and dynamic causal modelling allow the connections between regions to be inferred. Another example is that of adaptation (in which neurons will response less to the second presentation of the same

stimulus) which allows inference to be made about the activity of subpopulations of neurons. Chapter 4 extended the psychophysics study in chapter 3 by performing an fMRI study using similar stimuli. This allowed the ways in which the brain weights self-motion cues to be investigated. The study also allowed attempts to be made at decoding the associated signals.

The final aspect of navigation in this hierarchy is the continuous updating of our location and orientation within our environment through the use of self-motion cues. Several regions of the brain appear to be associated with these processes, with much of our knowledge based on animal models. For example, in rats, the head direction cells of the post-subiculum are vital for tracking orientation (Calton et al., 2003), while place-cell activity in the hippocampus is critical for navigation (Miller & Vogt, 1984). Chapter 5 investigates these more abstracted layers of spatial processing in humans, and looks at the representation of both place and orientation in several cortical areas. This chapter concludes the tracing of the self-motion hierarchy, from simple planar motion processing through its integration with other cues to the updating of orientation and place representations.

The hypotheses being tested in each chapter vary considerably. In chapter 2 the structure and organisation of hMT+ and V1 are investigated. In particular, the presence of columns is determined using a new method which looks at the correlation in the responses to opposing directions of visual motion stimuli. This method depends on neighbouring columns preferring opposing directions of motion, a feature found in optical imaging studies in primate MT. The anisotropic organisation of V1 was also revealed through the combining of multiple participants responses. Chapter 3 develops an audiovisual stimulus for the later fMRI study. In particular the study tests whether the visual and auditory motion cues can be integrated and, if so, by what rule they are combined. Several integration models are compared to see which one fits the data most successfully. These include Bayesian integration and a model of neural integration known as the Normalisation model. The later fMRI study will use adaptation, so posthoc adaptation analysis was performed on the results from the psychophysics study. Intriguingly, considerable behavioural adaptation was found. The two models (Bayesian and Normalisation) were modified to incorporate neural adaptation in a simple attempt to localise where the adaptation was taking place (either at an early unimodal or later integration stage). The modified model reproduced the results with adaptation localised to the unimodal layer of its design. This suggests that the behavioural adaptation we reported is due to neural adaptation in unimodal sensory regions, not necessarily in integration regions. A conflict condition was also included in the study to test whether the two stimuli are equally weighted, or whether one modality overrides the other. Visual capture was found to occur for many (but not all participants). To model this, a fitted prior was added to the Bayesian model which was able to both model the conflict condition and still successfully predict the results of the congruent condition.

Chapter 4 uses a variety of fMRI methods to investigate the neural responses to the same stimuli as that used in the psychophysics experiment. In particular we hypothesised that the weighting of two stimuli by their coherences is mediated by changes in the connectivity between the unimodal sensory regions (e.g. A1) and an integration region (which we hypothesise is VIP). Before testing this hypothesis, other predicted features were tested. The type of stimulus (visual or auditory) was successfully decoded in all regions, but the direction of the stimulus was only decoded reliably in V1 and (possibly) in A1.

This weak classification accuracy hinted that further problems would be experienced in other analyses. Dynamic Causal Modelling (DCM) was used to investigate how connectivity changes between the unimodal regions and the possible integration region (putative VIP) and was also used to test whether feedback connections were plausible as we would expect. Specifically, our key hypothesis was that the unimodal region associated with the most coherent stimulus would be the region most strongly coupled to the integration area. Our second hypothesis was that all feedforward connections are best modelled by being paired with feedback connections (based on many lines of evidence, see section 4.2.3 for more details). The model comparison results were found to strongly support this hypothesis, suggesting that the presence of feedback responses can be determined using these methods. However, our key model comparison (between a model with and without coherence modulation on its afferent connections) failed to properly find the modulation we had hypothesised. We also used Structural Equation Modelling to test a similar hypothesis, that in the high visual coherence conditions there would be a greater correlation between the visual cortex and the region pVIP (believed to be the integration region). Chapter 3, looking at models for the psychophysics experiments, includes a mechanistic neural model called the Normalisation Model, based on the work of Ohshiro et al. (2011). This model biases the inputs from the two unimodal regions by the reliability of the stimuli, and it is this changing weight which we would expect to see in the DCM or SEM results. A similar model (Ma et al., 2006) was considered which does not have this reweighting element. If the DCM results had been more reliable we may have been able to distinguish between these aspects of the two models.

The study also used adaptation analysis to test whether the putative VIP region contained cells which responded to both stimulus types. The *behavioural* adaptation results (in terms of the accuracy of responses) replicates the behavioural adaptation results of the earlier psychophysics study. However, the *fMRI adaptation* analysis of the direction of the stimulus motion, for both hMT+ and A1 do not appear to be significant (although direction adaptation was found to be significant across those subjects who were given a low-coherence version of the stimuli).

The study was not able to resolve its key hypothesis, that coherence modulates connectivity. The null result may be due to the absence of such a modulation, but the absence of other results (such as superadditivity and classification) suggests the stimuli require modification to improve the signal strength. The final chapter moves beyond egocentric cues into the realm of allocentric navigation. In particular it investigates the potential for fMRI to detect and decode the activity of head direction cells. Head direction cells, found in other species, each have a peak response in a particular preferred allocentric direction. The cells, in effect, act like a localised compass, allowing the animal to track its orientation in the environment. The relatively conserved nature of the hippocampal formation means that it is plausible that humans have such cells. It would be useful for future research in this field if the activity of such cells could be decoded from fMRI data. This follows from recent advances in fMRI in which activity in the hippocampus was able to predict the location of the participant in a virtual environment (Hassabis et al., 2009; Rodriguez, 2010). The hypothesis therefore is that we can decode the direction the participant is facing in the virtual environment from regions such as the posterior cingulate, the retrosplenial cortex and the thalamus, where head direction cells are predicted

to be found, and secondly we can decode the participant's location from activity in the hippocampus. The head direction however was not classified above chance from any of these locations, while the participant's place was only just decoded above chance (from the hippocampus), across participants. It was suggested this was due to differences in the organisation of neurons in the two regions, although limitations to the stimulus may also be a cause.

In summary, the study investigates the full hierarchy of self motion processing, from largely unimodal regions hMT+ and V1 to integration area VIP and finally allocentric processing in the hippocampal formation.

1.1 fMRI Methods

Each experimental chapter begins with a detailed background of the relevant fields. This is followed by detailed methods for that experiment. However, three of the four experiments which follow use functional Magnetic Resonance Imaging (fMRI), and as such share many methods and background information, which can be covered in one section. To this end, I will first introduce fMRI, its strengths and weaknesses, why it was chosen and what relationship it has with the underlying brain function we are trying to infer. The particular parameters used during imaging are also discussed.

1.1.1 fMRI: Background

Magnetic Resonance Imaging

Physical Basis Magnetic Resonance Imaging is an imaging technique which makes use of the alignment and precession of certain atomic nuclei. In the simplest case, hydrogen ^1H nuclei precess (their axes rotate) around the direction of an applied magnetic field[1]. This is known as Larmor precession and its angular frequency σ is proportional to the magnetic field strength B, by a factor γ, which is unqiue to each element. The direction of the precession axis can either be in the same (higher-energy state) or opposite (lower-energy state) direction to the magnetic field. When the nucleus switches from the high to low state, it emits electromagnetic (EM) radiation at the Larmor frequency. Similarly, nuclei can be moved from the low to the high energy state by receiving radiation at that frequency. At equilibrium, slightly more nuclei will be in the low-energy condition. This bias means that the vector sum of the magnetic field produced by all the nuclei, M_0, will lie in the direction of the applied field B_0. This component is known as M_z. Note that M does not have a component orthogonal to the applied field (M_x or M_y), because the nuclei are not precessing in phase (so the individual contributions to the magnetization cancel out). If an EM pulse is applied at the Larmor frequency (for most MRI this is a radio-frequency pulse, e.g. at 100MHz), a subset of nuclei will be moved into the higher-energy condition. The effect of this on the magnetisation is for the component in the direction of the applied field to be reduced ($M_z = 0$) or even reversed.

[1] This is the classical physics description of this phenomenon

This RF pulse is equivalent to an oscillating magnetic field, B_1. In MRI scanners, this is applied in an orthogonal direction to B_0, so that it causes the magnetisation vector to move into the M_{xy} plane. As this field oscillates at the Larmor frequency it emulates a field rotating at the Larmor frequency around the z-axis. The B_1 field rotates the magnetisation vector by the *flip angle*. For example 90° or 180° from the M_z direction.

After the RF pulse, the nuclei will immediately begin to go out of phase (spin-spin relaxation) and, they will fall back to the lower energy condition (spin-lattice relaxation). Both these processes of relaxation occur as forms of exponential decay. For the latter, the time constant T_1 describes the speed at which the M_z component returns to equilibrium. For the former, two effects are involved. The first (which causes the pure T_2 decay) is due to the energy being transferred between nuclei. The second effect is caused by the nuclei each experiencing slightly different magnetic field strengths. This means the Larmor frequency will vary slightly between nuclei, causing them to go out of phase. The local magnetic field strength varies because either the applied field is not homogenous or because of local interactions between atomic and sub-atomic particles. The effect of both is combined into the time constant T_2*, which describes the rate at which the M_{xy} component decays. Note that T_2* is always less than or equal to T_1. The rotation of the magnetisation vector around the z-axis causes the emission of EM-radiation at the Larmor frequency. This signal is called the free induction decay (FID), and is the signal received by the MRI scanner during imaging.

Flip Angle selection With a very long TR, one wants to maximise the component in the M_{xy} plane, and so a flip-angle of approximately 90° is used. For short TRs however, when the slice will be repeatedly scanned, the spins will not have completely relaxed to be in the M_z direction before the next RF pulse. To take this into account the flip-angle needs to be reduced.

If the tissue has a spin-lattice relaxation time constant of T_1, then for a given TR the proportion of the signal remaining at the end of the TR will be $^{-TR/T1}$, e.g. if TR=200ms, T1=1400ms, $^{-200/1400} = 0.867$, only a small portion of the signal will have decayed by the end of the TR. This equates to a reduction in the angle of the magnetisation vector of $arccos(0.867) = 30 degrees$. To recover this lost angle for a second scan requires the RF pulse to induce a flip angle of 30°. A larger or smaller value would mean the vector would have a smaller M_{xy} component, thus weakening the signal.

Imaging To perform imaging, one has to be able to localise where a particular signal is from.

Slice Selection: One first restricts the imaging to a single slice through a process called slice selection. This is acheived by making use of the relationship between field strength and Larmor frequency. By applying a gradient to the magnetic field, the Larmor frequency varies across the brain. The RF excitation pulse will only match the frequency of a small slice of the cortex, which means only that slice will be excited. A carefully shaped (sinc-function, in frequency-space) signal is used to make the slice 'square' and of a precise thickness. We assume here that the slices are normal to the z-axis.

Frequency encoding: If one imagines that the slice contains two small spin-containing regions (e.g. two glasses of water), we can use a gradient in field strength to distinguish them. By applying a gradient

(orthogonal to the slice-selection gradient) in which the field is stronger in one glass than another, we will receive two distinct frequencies from the sample (due to the two different Larmor frequencies induced by the differing field strengths). One can infer the location and quantity of spin-nuclei across that dimension by looking at the frequencies of the received EM radiation. This is extracted from the received signal by applying a fourier transform. In this example we assume this *frequency encoding* is along the x-axis.

Phase encoding: To allow inference about the spin-density in the y-axis, we can make use of the phase of the spins. By briefly applying a gradient to the magnetic field, across the y-axis, we are able to change the phase of the nuclei in that direction. We therefore have two parameters, the frequency and the phase of the spins. This provides a unique pair of values describing the location of any particular spin. However, in general, it is not possible to uniquely determine their locations from just one phase/frequency signal. To solve this the sequence must be repeated with a different phase encoding gradient. This allows unique lines to be recorded in the frequency/phase gradient domain (referred to as k-space), each for a different phase-encoding gradient amplitude. The points along the line record the response for different frequency encodings, and the fourier transform allows the experimenter to generate the image from the k-space image.

Echo-planar Imaging: The above procedure requires one line of k-space to be recorded at a time, by reapplying the phase-encoding gradient. So to image a single slice requires many (e.g. 128) TRs to be repeated. Instead, to image a whole slice quickly, one can use a technique called *echo-planar imaging*. After the initial frequency and phase encoding, a 180° RF pulse is applied. This is followed by a series of frequency encoding pulses and phase encoding pulses, which effectively allow the whole of k-space to be traversed in a single TR.

Structural MRI often uses spin-echo imaging. By applying a 180° pulse the loss of phase coherence is cancelled out (as the phases are rotated in the opposite direction after the ‘refocusing’). Note that this cannot be used for fMRI as it is the T_2* relaxation which relates to the proportion of oxygenated haemogolobin in the voxel.

functional MRI As mentioned above, T2* decay describes how quickly nuclei in a voxel lose coherence (go out of phase). Deoxygenated haemoglobin (dHb) is more magnetic than when oxygenated (Hb). The diamagnetic deoxygenated haemoglobin alters the local magnetic susceptibility (how magnetised the material becomes in response to the applied magnetic field). This causes tiny inhomogeneities in the local field and thus the rate at which the nuclei go out of phase (spin-spin relaxation).

During fMRI the general sequence of events begins with an increase in neural activity leading to greater metabolic activity and an associated increase in the rate oxygen is absorbed from the surround blood vessels. However, this is more than compensated for by an increase in the blood supply provided by the dilation of the surrounding vessels. This increase in Hb means the spin-spin relaxation becomes less, which is then detected as an increase in the T_2* weighted BOLD signal during fMRI.

The mechanism which causes the dilation of local blood vessels is currently still poorly understood. It may be mediated by glial cells, which respond to neural activity by increasing the concentration of

intracellular Calcium ions (Schipke & Kettenmann, 2004). This in turn causes the release of further signalling chemicals, including nitric oxide and epoxyeicosatrienoic acid (Metea & Newman, 2006) which then causes the dilation of blood vessels. This effect could propagate over considerable distances, mediated through glial Ca^{2+} waves (Schipke & Kettenmann, 2004). Intriguingly the review notes that neural activity can induce these waves but also that the waves were reported to have an effect on neural activity.

1.1.2 Neurovascular coupling

The full link between the prevalence of oxygenated Haemoglobin (Hb) and neural activity is complex and far from fully understood. A study by Logothetis et al. (2001) investigated by recording local field potentials (LFP), single and multi-unit activity whilst simultaneously using an EPI sequence to measure the associated T_2*-weighted BOLD signal, from the visual cortex of monkeys. LFPs relate more to subthreshold integration within neurons, than to local spiking activity. These signals are generated by the input from more distant regions, and so relate to cortex within a millimetre or so of the recording site (Logothetis et al., 2001). They found that the LFP in the voxel's vicinity was the most correlated with its BOLD response (compared to the unit recordings), suggesting that the BOLD contrast is correlated most strongly with a region's inputs (and processing) than its outputs. A later follow-up study in awake monkeys confirmed this earlier finding. Both these studies only looked at this coupling in the primary visual cortex. It may be that some features of this region of the brain make the LFP a particularly strong correlate of the BOLD response. Further studies are probably required to understand neurovascular coupling.

Other researchers (e.g. Buxton et al., 2004) have attempted to create generative models to link the neural activity to the BOLD response. These follow the pattern of stimulus and neural response, which is then used to predict the cerebral blood flow and rate of oxygen use. These are entered into a model of how the veins respond to changes in blood flow, called the balloon model. This is then used to predict the blood volume and the proportion of deoxygenated haemogolobin. These predictions are combined to calculate the modelled BOLD response. Such generative models are also the basis of Dynamic Causal Modelling, an attempt to infer changes in connectivity and activity using the BOLD response (Friston et al., 2003).

1.1.3 fMRI: Strengths and Weaknesses

Functional imaging has come to dominate the field of cognitive neuroscience, with tens of thousands of papers published each year using the technique. Its popularity over other types of imaging and recording tools is almost certainly due to its non-invasive nature. It doesn't require surgery, ionising radiation or the injection of dyes or drugs. This makes it ideal for use with human participants. Humans, besides the advantage of being able to consent to take part and being of particular interest to researchers, can also be instructed to perform particular tasks whilst in the scanner. Compared to some other non-invasive imaging (such as Electroencephalography, EEG and Magnetoencephalography, MEG) the spatial resolution

of fMRI is both very high and unlike EEG and MEG the signal source is not restricted anatomically. The resolution of MRI can reach the $100\mu m$ scale (Bolan et al., 2006), at which the size of the haemodynamic response becomes the fundamental limit. The Bolan study of the cat visual cortex achieved such resolutions by using a narrow bore magnet, a contrast agent and a general anaesthetic, combined with a very high field strength (9.4T). Human fMRI imaging is now also possible at sub-millimetre scales using 7T fMRI scanners and a bite-bar to minimise head motion. For example Yacoub et al. (2008) used $3 \times 0.5 \times 0.5mm$ voxels when imaging orientation columns. The resolution clearly however can not compete with single- or multi-unit recordings performed during non-human or in-vitro experiments. These allow the responses of, for example, individual place and head-direction cells to be detected. Besides the limited spatial resolution, fMRI suffers from an intrinsic limitation in its temporal response, which means many questions one might hope to answer with fMRI may be intractable. For example, determining the direction of connectivity using fMRI is largely impossible due to the latent and sluggish BOLD response. The bidirectional nature of much of the brain's organisation makes inferences about connectivity and even the localisation of particular processes particularly difficult. The indirect nature of the BOLD response, besides causing this temporal inaccuracy, also causes the response to vary between regions and conditions, with the added concern that there is considerable uncertainty about which neural activity would cause any given response.

Many methods have been developed to try and utilise the fMRI data as effectively as possible. Two used in this study are multivoxel pattern analysis, which combines the data from many voxels to extract evidence of a signal of interest, and the second is the use of adaptation analysis, in which the experimenter attempts to modulate the response of a subpopulation of neurons within a voxel to allow inferences to be made about the preferences of the subpopulation.

Other limitations include: Limits on the extent to which one can compare different brain regions (due to differences in the way information is processed and in the BOLD response to equivalent signals); The use of the subtraction paradigm, which assumes linear interactions between the conditions being controlled; And assumptions about what the baseline is when contrasting conditions - the brain is never completely 'at rest'.

Finally, the limited spatial resolution and its indirect nature mean that the fMRI BOLD signal will always be reflecting the responses of a massive set of neural (and other) activity (Logothetis, 2008), leading to the conclusion by Logothetis that fMRI will only ever be a supporting tool, for understanding brain function. While potentially overstating the case, progress is often greatest when the results from optical imaging, single-cell recording, anatomical studies and fMRI are combined.

1.1.4 Choice of Protocol

The imaging for the self-motion and head-direction chapters was conducted on the Siemens 3T TIM Trio, using a 32-channel head coil. For the experiment in chapter 2, the images were acquired using a GE Sigma HDx 3T scanner with an 8 channel head coil.

Initial localiser Shimming was only applied once, at the very start of scanning. Participants were asked to keep their heads completely stationary during the whole scan. Head movement during the scan could cause this shimming to become less effective. Head motion was inspected offline and only one participant was rejected due to large and repeated head-motion.

All the MRI scans began with a 6 slice 2-axis anatomical localiser, to allow the placement of the scanned volume. In the self-motion study (Chapter 4), this was followed by the full anatomical scan while in the other studies it was followed by the functional imaging, with the T1-weighted structural scan at the end of the experiment.

For some participants, the radiologist at the CCNi made use of the AAScout protocol to assist in quickly aligning the scanned volume between subjects, however the volume was always manually adjusted prior to scanning to ensure coverage of the regions of interest.

T1-weighted structural scan The anatomical scan in all three fMRI experiments was a standard Siemens Magnetization Prepared Rapid Acquisition Gradient Echo (MPRAGE) scan, with $1mm^3$ isotropic voxels. TR=1900ms, TE=2.52ms, FA=9°, using two-fold iPAT (a method for reducing scan time by combining the gradient encoding of spatial information with the signal from the RF coil). The scan took approximately 6 minutes.

T2*-weighted fMRI scan The final fMRI chapter (head direction) used a TR of 1940ms, a TE of 30ms (this combination generates a T_2* weighted image, as both the TR and TE are relatively long). Based on T_1 = 1300ms, a flip angle of 77 degrees is used, based on the Ernst Angle computed by $arccos(e^{TR/T1})$ substituting in the values of TR and T1, $arccos(e^{-1940/1300})$. The imaging matrix size was 100 × 100 covering a FOV 200mm on each side. This leads to a voxel size of $2mm \times 2mm \times 2mm$. The whole volume was made with 30 slices, each 2mm thick with a 10% gap. The iPAT factor was 2 and the EPI was single shot (no interleaved sampling of k-space). To minimise cross-talk between slices, the order of slices was interleaved, and slice timing correction applied using the SPM 8 toolkit. In retrospect, interleaving appears to be a largely unnecessary step, as the gap and the precision of modern RF excitation pulses largely avoids cross-talk[2].

The self-motion experiment in chapter 4 used a TR of 2200ms, and a TE=30ms. The flip angle chosen was the same, at 77 degrees. The image matrix was also 100 x 100 with an identical field of view (200mm), so the voxels were $2mm \times 2mm \times 2mm$ in this experiment also. 35 slices were taken in each volume, leading to the slightly longer TR. To avoid extending the TR to much further, a 25% gap was incorporated. The default iPAT value of 2 was chosen as a balance between the slight signal loss and the increase in sampling rate it affords.

In the experiment in chapter two, the imaging sequence used was also the standard echo-planar sequence (TR: 2000ms, echo time: 40 ms, flip angle: 70° chosen with a slightly different assumption about the value of T_1). There were 25 slices per volume with interleaved acquisition and no gap. Voxels

[2]From 'Operating the Scanner: Should I use interleaved slices?' downloaded from Harvard University: Center for Brain Science. Last accessed: July, 2013

were $1.75mm \times 1.75mm \times 2.4mm$ in size.

PACE, MOCO and other online motion correction methods are not used. Instead the motion correction is restricted to offline processing only, using the SPM toolkit. This involved motion correction (realignment) of all the fMRI volumes to the first of the fMRI images. This step also uses SPM's unwarp function which attempts to correct for nonlinear distortions caused by the inhomogeneities in the magnetic field, which change as the subject moves. Note, besides the two 'dummy-scans' of the Siemens standard BOLD sequence, several volumes were effectively discarded as there were several seconds between the initial pulse received and the start of the stimulus.

Resolution Selection Unlike a standard 'GLM' fMRI study, it was not planned to spatially smooth the voxel data (a standard step during GLM analysis). Instead, the data was being used in such a way that small scale differences between voxels were believed to be important for the success of the classifier, although some evidence suggests that it is larger scale differences in activity which allow the classifier to function (Freeman et al., 2011). It is worth noting here that although spatial smoothing may not greatly harm classification (Kamitani & Sawahata, 2010), a reduction in voxel resolution is not equivalent to spatial smoothing, as it consists of an uninvertible down-sampling across space. An increase in resolution comes at a significant cost in a reduced signal to noise ratio. This is discussed in more detail in later chapters.

1.1.5 Artefacts

Common Artefacts Several different artefacts can be found in MRI images. These include:

Ghosting: This is usually caused by the movement of the subject, which will cause changes in the field inhomogeneity. This means that the frequency will not be quite the same across the slice leading to changes in phase. For example if the frequency falls from 100Mhz by 50Hz, after 2ms that region will be out of phase by 10%. This causes an ambiguity in the location of the spins in the phase direction, hence why ghosting appears generally in the phase-encoding direction. In retrospect some aspects of the slice prescription should have been considered more carefully. For example in the experiment in chapter 5 the volume was placed so that important slices included the subject's eyes. This could have been avoided so that motion artefacts from eye movement were minimised. It is worth noting that this is in the frequency-encoding direction and so is of less concern.

Susceptibility: The susceptibility of a material is the degree to which a material can be magnetised, in response to an applied magnetic field. Where large changes in the susceptibility of materials occur (e.g. at air-tissue boundaries) the local magnetic field changes in a very non-linear manner. This can't be completely corrected through shimming and means the images become distorted and often appear to have reduced signal. Particularly for the study looking at the hippocampus, signal drop out (due to the proximity of the hippocampal region to the sinuses) was of concern. The use of the newly installed 32-channel head coil appeared to largely mitigate the effect. In retrospect it may have been desirable to maximise the SNR from the region by using passive shims (Wilson et al., 2002), or other methods to

avoid signal loss in that region.

Shape distortion: Even with shimming the magnetic field across the human head will vary by 1ppm, this leads to a slight displacement between the anatomical and functional images depending on the magnetic field strength across the brain. This becomes an issue as we generally want to align the EPI functional images with the T_1 structural scans. Software is used to coregister these images. In the SPM8 toolkit used during this PhD, this involves rotation, skew, and scaling transformations to align the two types of image.

Phase-wrap: Aliasing, caused by the sampling being only within the field-of-view. Anything outside the FOV will be wrapped to the opposite side, as it will be responding with the same phase (the phase obviously repeats outside the FOV). The phase direction must generally be chosen to avoid objects outside the FOV. These problems can be largely mitigated in the frequency direction by oversampling.

Protocol To look for problems in the collected data the Artefact Repair Toolbox (v4) SPM8 tool was used to allow both a visual and an automated inspection to look for failed volumes. The difference between consecutive volumes was used as this revealed motion artefacts, ghosts and other problems most clearly. The data from almost all participants was considered of sufficiently good quality. However there were two participants scanned during the head-direction experiment (Chapter 5) which were discarded completely from further analysis due to multiple volumes being badly affected by waves across multiple slices and volumes (similar to those one would expect to see if an RF spike was introduced into a single pixel of k-space). It was found the scanner needed repair prior to further imaging.

Summary Although the imaging was carried out with sufficient precision, in retrospect there were improvements to the protocol that may have further improved the SNR, such as the choice of slice orientation, voxel size and stimulus coherence.

Chapter 2

Columns and Maps in hMT+ and V1: Responses to Simple Planar Motion

---

The Chapter below is based on the paper *in submission* (Smith et al., 2013). The stimulus and fMRI scanning were designed and performed by Magdalena G. Wutte and her supervisors Virginia L. Flanagin and Stefan Glasauer, for the purposes of another experiment (Wutte et al., 2011). Their research was supported by the Deutsche Forschungsgemeinschaft (GRK1091, JA1087/1-1). I was asked to assist in the analysis of the data for that earlier study, and the following findings were incidental to that analysis. The chapter is supplemented with some additional related findings, not included in either paper.

Prior to investigating the integration of cues, we investigated a purely visual stimulus. Congruent visual motion (optic flow) is a particularly salient self motion cue (Lappe et al., 1999). For example, Prokop et al. (1997) found optic flow had a strong influence on stride length, while Lee & Aronson (1974) found that invalid optic flow cues could cause human infants to fall over. Before experimenting with a complete optic-flow field, we explored the organisation and anisotropies of the human visual system for simple linear visual motion. Although not highly naturalistic, this stimulus provides a useful way of investigating the visual system's organisation. Importantly the anisotropies revealed by this experiment in V1 may be directly linked to the anisotropies in the visual flow fields which humans experience while navigating in a normal environment. The research focuses on the columnar organisation of hMT+ and how it might be detected using fMRI.

Cortical columns are found in many regions of the brain and are a common structure in the organisation of neural preferences. In humans, evidence for such organisation is limited by the resolution of imaging methods. Here, we describe a novel method correlating the BOLD signal response to different pairs of visual motion stimuli. In the human motion complex, hMT+, we observed that BOLD signal responses to opposing directions of motion were more correlated than the responses to orthogonal

directions. In contrast, the same analysis on V1, a region also found to contain direction information but believed not to contain direction columns, did not reveal correlations between BOLD responses to opposing directions. A likely explanation for these findings is that it is the columnar organisation of hMT+ which causes the correlation effect, in which cortical columns with opposing direction preferences are paired together. Our correlation method could therefore be a valuable tool for exploring columnar organisation of motion processing in the visual system.

2.1 Introduction

Figure 2.1: The optic flow field as one travels along a linear trajectory over a landscape. The arrows indicate the local optic flow. The circle indicates the current heading direction which, in this case, is also the same as the *centre of expansion*. Photo taken on the Isle of Coll

Optic flow is the pattern of motion on the retina which occurs when an organism moves (Gibson, 1950). This optic flow field depends on the distance or depth of the scene and the organism's speed and direction of motion. Figure 2.1 illustrates the optic flow one might perceive when moving in a straight line across a landscape. When travelling in a straight line with a fixed gaze, the centre of expansion lies in the direction of the animal's heading. However, if one tracks a landmark or moving target, the optic flow field on the retina will no longer follow this simple rule and will instead depend on the additional complication of the rotation rate and direction of the eye. This is known as the 'rotation problem', and requires the visual system to separate the visual motion caused by the movement of the eye from that caused by the movement of the organism.

It is a widely accept model of the visual system that it consists of two separate processing streams, one processing motion and location and the other processing identity and recognition. These are referred to as the dorsal and ventral streams, respectively. This distinction may extend to the frame of reference, with the dorsal stream appearing to remain in largely egocentric coordinates. Key features of dorsal stream processing are the many regions containing neurons which have been found to be particularly responsive to visual motion, such as MT, MSTl/d and VIP. Note that in this study the regions MT and MST are not distinguished. Both MST and VIP (both downstream from MT, figure 4.1a) have much

larger receptive fields than MT making them particularly suitable for calculations attempting to process the optic flow stimuli. In particular the motion patterns MST and VIP neurons respond to include complex expansion and rotation stimuli (Duffy & Wurtz, 1991). A final, highly relevant feature of MST is that it responds to vestibular cues (Fetsch et al., 2007; Gu et al., 2006); this is a strong indicator that the region is of importance in processing self motion stimuli. In this study we focus purely on linear 2d motion, with minimal eye or head movement. We also unfortunately are required to combine MT and MST due to an absence of anatomical markers for the regions, and the lack of a functional localiser used during the experiment.

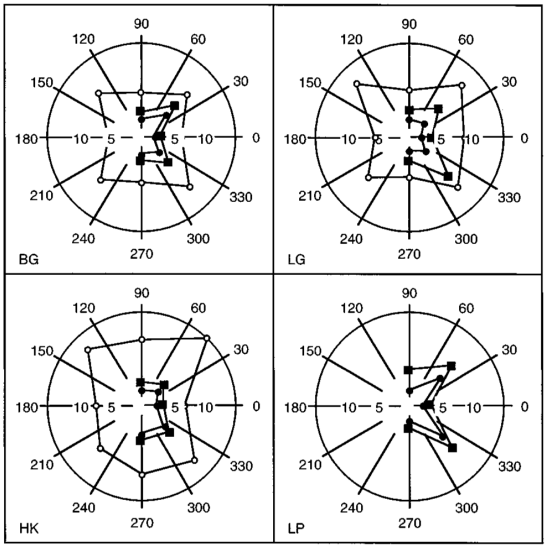


Figure 2.2: Discrimination coherence thresholds for different reference directions, from Gros et al. (1998), figure 5 (open circles, 25% coherence; squares, 50%; filled circles, 100% coherence). It is clear that oblique directions have the worst (highest) thresholds, but it appears the vertical directions have slightly worse thresholds than horizontal.

In addition to the main hypothesis regarding columnar organisation we considered whether we would be able to detect the physiological consequence of a form of the oblique effect, but for motion. For gratings partly masked by white-noise, vertical and horizontal gratings are most easily detected. Oddly though, when partly masked with natural or pink-noise the oblique gratings are more easily detected than the horizontal or vertical lines (Hansen & Essock, 2006), with horizontal grating perceived the least clearly. Hansen & Essock (2006) propose that the reduced perception at those angles is due to the inhibition in the visual system which causes the more represented orientations to become more inhibited. This allows novel content to be more easily detected. This experiment differed from others such as Girshick et al. (2011) in that it describes a detection task, rather than an descrimination task.

An important effect is the width of the tuning curves of individual neurons varies depending on the orientation, maybe also due to the increase in inhibition around those directions (Li et al., 2003).

A few studies have investigated anisotropies in motion detection and distinction, although most look at the oblique effect rather than the horizontal/vertical difference. Gros et al. (1998) suggest that motion sensitivity is isotropic while motion discrimination is better in the cardinal directions (anisotropic). Our experiment was slightly different, with the discrimination task being between speed and not direction. Gros et al. (1998) did not report any significance in discriminating direction, between horizontal and vertical directions of motion, however Figure 2.2 (from that study) suggests there might be slightly worse direction-discrimination thresholds for vertical motion. An interesting possible interpretation is that it is the horizontal speed of the dots that was poorly discriminated, suggesting that vertical speed discrimination is better than horizontal.

A final consideration is that, with a greater number of horizontal edges in natural scenes (see supplementary figure 4 Girshick et al., 2011), vertical motion would be more common, due to the restriction on perception of motion to directions orthogonal to edge orientation. A hypothesis for the distinction between motion speeds in the two orientations can also be proposed. Following the logic of Hansen & Essock (2006) and Girshick et al. (2011), with vertical motion possibly being more common we would expect motion in that direction to be more represented in the cortex. Note that although these two papers appear to contradict each other, both agree that more common stimuli are likely to be better represented in the cortex, and are likely to have better discrimination thresholds. Hansen & Essock (2006) also suggest that the detection of these stimuli might be more challenging due to the possible role of inhibition.

Functional optical imaging of cortex bridges the gap between low-resolution fMRI and small-scale electrophysiology, allowing us to see many structures difficult to infer using the other methods. For this study the most relevant optical imaging result is that of Kaskan et al. (2010), who investigated the organisation of direction-of-motion columns in MT in owl monkeys. In particular I was drawn to a figure in their paper (Figure 2.3) which shows the responses to drifting gratings in 8 directions, without subtracting the opposing directions (a processing step usually applied when publishing such images). The figure appeared to show the neighbouring columns were for opposing directions. We test this hypothesis using these images as it has such importance to our study.

Cortical columnar organisation has been found across species and brain regions (Hubel & Wiesel, 1968, 1974; Mountcastle et al., 1957), and appears to be a fundamental organisational principle across sensory and polymodal association cortices. In humans, neuroimaging is currently the only non-invasive method available for investigating functional columns. High field, high-resolution functional Magnetic Resonance Imaging (fMRI) has been used to detect ocular dominance columns (Cheng et al., 2001) and orientation columns (Yacoub et al., 2008) in human primary visual cortex (V1) and a recent study using high field strengths has hinted at direction columns in hMT+ (Zimmermann et al., 2011). However, such experiments are limited by the need for very high field strengths, good resolution, preferably flat cortical areas, new imaging methods and large Blood Oxygen Level Dependent (BOLD) responses. In particular to find columns in regions with complex cortical folding, the imaging must be 3d and have high resolution both in-plane and in slice-thickness. They extrapolated from data from monkeys that one cycle of the human direction column map may be of the order of 2 to 2.8mm. For reference they

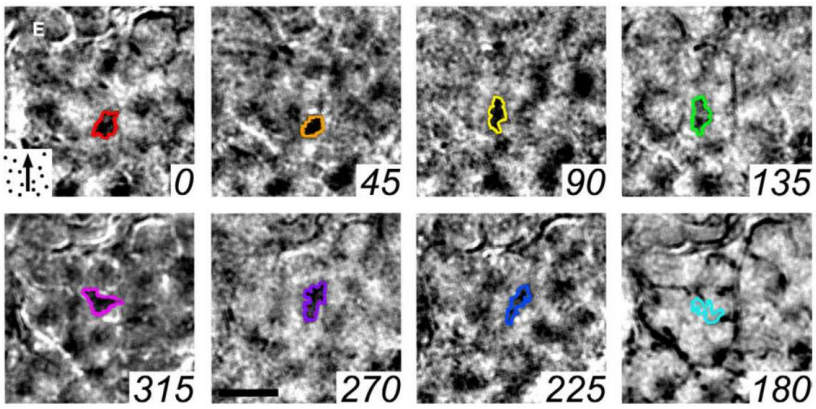


Figure 2.3: Optical imaging direction maps from Kaskan et al. (2010), figure 3. These show the responses to eight directions of dot motion. The same, averaged, cocktail blank was subtracted from all eight. Although difficult to determine by-eye columns of opposing directions appeared to neighbour more than one would expect by chance. Figure 2.7 in the results section reveals this feature more clearly. The scale bar is 1mm. To give an idea of scale, note that roughly three of this study's fMRI voxels will fit into one these images.

imaged with a voxel resolution of $0.8mm$, which meant resolving the columns required a regular pattern in the map to allow the sampling to detect the cycle of columnar preferences.

In this chapter, we present an alternative method for detecting columnar organisation, using fMRI data acquired with standard spatial resolutions ($1.75 \times 1.75 \times 2.4$ mm voxels) and standard field strengths (3 Tesla). We use as our region of interest an area of the brain heavily involved with motion processing. Animal studies have found region V5/MT to be receptive to visual motion and to have columnar organisation (Weliky et al., 1996; Xu et al., 2004; Kaskan et al., 2010). fMRI studies have found that the human Middle Temporal complex (hMT+, which probably contains the human homologue of V5/MT) responds strongly to visual motion, contains direction-of-motion information (Kamitani & Tong, 2006, e.g.) and may have columnar organisation (Zimmermann et al., 2011). To test the hypothesis that humans have columnar organisation in hMT+, we exploited the expected organisation of direction preference columns. Both electrophysiological (Albright, 1984) and optical imaging studies (Malonek et al., 1994) in monkeys have found that adjacent direction preference columns prefer opposite directions of motion. If a similar organisation exists in parts of human hMT+, one would expect the cortex within an fMRI voxel to contain more columns from two opposite directions, and fewer from other (orthogonal) directions.

To test our hypothesis that humans have columns in hMT+, we presented participants with dot-motion stimuli in one of the four cardinal directions while measuring the induced BOLD signal in two visual regions of interest (ROI) using fMRI. We predicted that the mean BOLD signal would be greatest for each voxel when the stimulus was of motion in either of two opposing directions (see Figure 2.4A for an illustration of upward and downward directions). When the results from multiple voxels are

combined (Figure 2.4B), one would expect a greater correlation between opposing directions (Figure 2.4C) than between orthogonal directions (Figure 2.4D). If found, this effect could be a marker for underlying columnar organisation.

The same analysis was performed on data from V1. V1 BOLD responses also provide information about the direction of visual motion stimuli (Kamitani & Tong, 2006), suggesting some form of direction-sensitive organisation. As this has been related to a large scale map rather than cortical columns (Freeman et al., 2011), we expected to find no correlation between responses to opposing direction in V1.

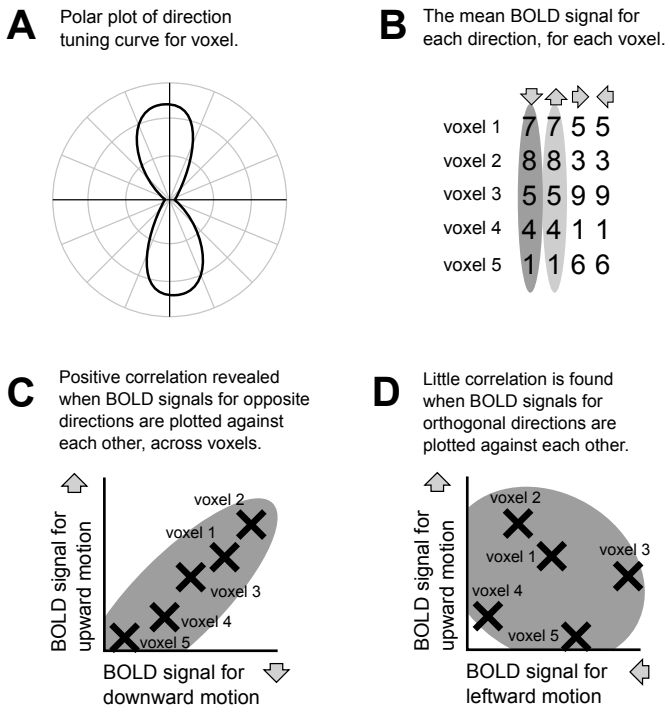


Figure 2.4: A. A fictitious voxel's BOLD signal, given different direction stimuli. B. An example table showing fictitious mean BOLD signals for each of the four cardinal directions, for each of the voxels in an ROI. C. The mean BOLD signal elicited by upward motion is plotted against the BOLD signal generated by downward motion, for each voxel. The result is a positive correlation. D. A similar plot between a pair of orthogonal directions results in an uncorrelated distribution. Note that partial correlations are used in the real analysis, to allow the voxel's baseline signal to be controlled for.

2.2 Materials and Methods

2.2.1 Participants

32 subjects (13 female) participated in this study. Each gave written informed consent to participate and understood the instructions without difficulty. The study was performed in accordance with the Declaration of Helsinki and was approved by the ethics committee of the medical faculty of the Ludwig-Maximilians University Munich. The subjects were recruited in two age groups (21-27 years and 60-72 years), for the purposes of another experiment (no group-effect in the main correlation result was found, $p = 0.64$). Two subjects were excluded from the analysis, one due to excessive head motion, and one due to fatigue during the fMRI scan, leaving a cohort of 30 subjects (12 female). All participants had sufficient visual acuity for the stimulus at 60cm distance (Snellen fraction was at least 0.4).

2.2.2 Magnetic Resonance Imaging

Experimental Stimulus

Visual stimuli were projected with an LCD projector onto a screen placed behind the participant. This was viewed using a mirror placed above the participant, angled at 45°. The dot motion stimuli were produced using Vizard 3.0 (Worldviz, www.worldviz.com). Coherent, translational flow fields were presented, within a circular aperture, with a diameter of 11.4°. A constant 300 dots were on the display at any time, and each dot had a diameter of 0.11°. The flow fields moved in one of four possible directions (Up, Right, Down or Left). Participants were instructed to fixate upon a cross in the centre of the circle.

Experimental Design and Task

Stimuli were presented in blocks, each lasting 18 s, during which the dot motion direction was kept constant. Each block consisted of four trials, and each trial consisted of 2 stimuli, one of which had slightly faster moving dots than the reference stimulus, which was fixed at $8°/s$. Each trial lasted 3.25 s, consisting of a pair of stimuli 1.5 s long, separated by a 0.25 s interstimulus interval. An intertrial interval of 1.25 s occurred after each trial. Each 18 s block was followed by a 10 s rest period (Figure 2.5C).

The subjects were given the task of reporting which was the faster stimulus within each trial by pressing one of two buttons. To ensure that the task difficultly remained constant, the speed difference between the reference ($8°/s$) stimulus and the faster stimulus was adjusted to keep the task performance at approximately 80% correct. The speed was set using an adaptive procedure and the QUEST algorithm (Watson & Pelli, 1983). This task was independent of the direction of motion condition.

All subjects performed 8 runs (except two subjects, who performed only 7 runs). Each run consisted of 16 blocks. There were a total of 32 repetitions per direction. The participants all practised the task outside the scanner, until they reached a specified performance level of 2 consecutive runs in which

at least 12 of the 16 blocks were error free, at which point they were required to maintain a stable 80% correct threshold for at least 12 trials, on a staircase procedure. Before the imaging started the participants practised inside the scanner.

Functional Localiser

A separate fMRI experiment was conducted to localise the hMT+ region in each subject, using previously established procedures (Huk et al., 2002; Morrone et al., 2000). Alternating stimuli of moving or stationary dots were presented for 16 s, in a circular aperture (11.4° diameter), separated by rest periods of 10 s. Moving dots travelled towards and away from the fixation cross at $17.1^{\circ}/s$. During the rest periods the screen was blank, except for a fixation cross. Subjects fixated at all times on the cross in the centre of the circle.

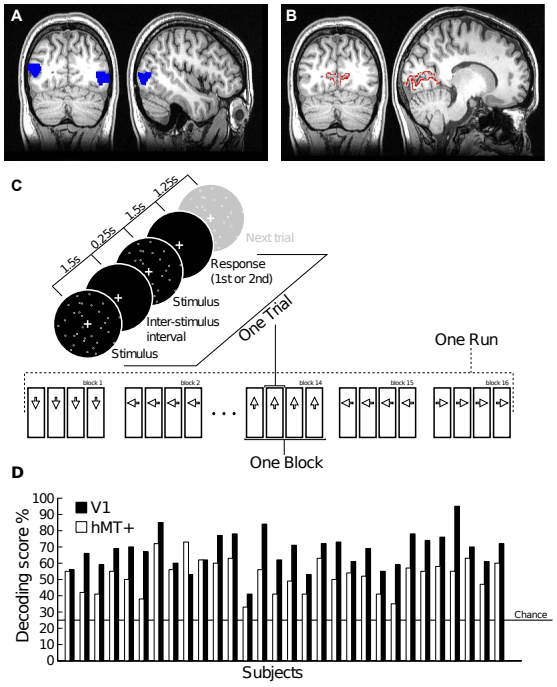


Figure 2.5: A. Coronal and Sagittal sections showing the functionally localised ROI for hMT+. Subject s01. B. Coronal and Sagittal sections showing the anatomically localised ROI for V1. Subject s01. C. Experiment design. The experiment was divided into eight runs. In each run there were sixteen blocks. The four trials within one block all have the same stimulus direction. One trial consisted of two stimuli (each 1.5s long), separated by an interstimulus interval (of 0.25s). The subject then had 1.5s to respond to which stimulus had dots moving more quickly. After an intertrial interval of 1.25s, the next trial started. D. The decoding scores for each participant using either the V1 ROI or the hMT+ ROI. Chance was at 25%. All subjects decoding scores were above 34% ($p < 0.01$), except for one subject in the hMT+ region. The results using the V1 ROI have a higher accuracy than those from hMT+.

fMRI Acquisition

The images were acquired using a 3T MR-Scanner (GE Sigma HDx) with an 8 channel head coil. The imaging sequence used was the standard echo-planar sequence (TR: 2s, echo time: 40 ms, flip angle: 70°). There were 25 slices per volume (interleaved acquisition, no gap), centred on the area of interest (medial temporal lobe). The voxels were $1.75 \times 1.75 \times 2.4mm$ in size. Each experimental run sampled 225 volumes, while the hMT+ functional localiser sampled 132 volumes. Finally, a T1-weighted anatomical volume was acquired.

2.2.3 Regions of Interest

Defining the hMT+ ROI

To define the hMT+ region of interest, fMRI data from the functional localiser was realigned to the first volume of the time series and smoothed with a 4mm FWHM kernel (SPM8, Wellcome Department of Imaging Neuroscience, London, UK). The images were kept in individual subject space, to avoid artefacts caused by spatial normalising. A general linear model (GLM) was used to find which voxels had an increased response to motion. The design matrix contained regressors describing the periods of dot-motion and periods of stationary dots. A contrast between motion and stationary regressors identified clusters for hMT+ bilaterally (FWE, $p < 0.05$ in all but three subjects, who showed hMT+ clusters only at $p < 0.001$ uncorrected, Figure 2.5A). The clusters from the two hemispheres were combined to make the hMT+ masks. Appendix D looks at the size of the hMT+ mask in more detail.

Defining the V1 ROI

Unlike the hMT+ mask, the V1 mask was defined based on anatomical landmarks. Figure 2.5B shows the initial mask which was created using the cortical parcellation algorithm provided by FreeSurfer (Fischl et al., 2002, http://surfer.nmr.mgh.harvard.edu) using anatomical constraints described by Hinds et al. (2008). As with the hMT+ mask, the two hemispheres were combined.

Multivoxel pattern analysis (MVPA)

We used the Princeton Multi-Voxel Pattern Analysis toolbox (MVPA, www.pni.princeton.edu/mvpa) to test whether hMT+ and V1 contain information about the direction of the stimulus. The fMRI data was preprocessed by realigning (SPM8) and high-pass filtering (MVPA, cut-off: 128s). Each time course was z-scored within runs to remove baseline differences between runs. The haemodynamic latency was corrected for by the standard practice of shifting forward the onset times by 4 s (Kamitani & Tong, 2006). Note that no smoothing or spatial normalisation took place. The 160 most significant voxels from the localiser's t-image were used to select the voxels from the hMT+ mask. This number was chosen based on previous work which found that the classification accuracy in hMT+ was not greatly improved if further voxels were added (Wutte et al., 2011). The same number of voxels were selected from the V1

mask to allow the two regions to be compared. The voxels were chosen without constraining how many came from each hemisphere to make up the 160. In summary, both ROIs were defined functionally using the results of the visual-motion localiser.

A linear support vector machine (with a fixed cost of one) was used to perform the decoding (LibSVM, www.csie.ntu.edu.tw/˜cjlin/libsvm). The time points within each block were averaged to give one value for each block. Classification used standard leave-one-(run)-out cross validation. The data was split, with all but one run being used to train the classifier, and the remaining run used to test the classifier's accuracy. The process was then repeated, until all runs had been used as the test set. A classification score was calculated, reporting the percentage of blocks in which the classifier correctly decoded the direction of motion.

2.2.4 Correlation

A separate analysis was performed on the same data, to investigate the correlation of responses to different directions of motion. The aim was to find the correlation in the responses to different pairs of directions, across the voxels in the ROI. This analysis also used just the 160 most significant voxels from the localiser's t-image, for both the hMT+ and V1 ROIs. We did not investigate which subfield of hMT+ the voxels were from.

Data Preprocessing

The fMRI data was realigned, z-scored (MVPA), and high-pass filtered (MVPA), as described for the pattern analysis. The values for each voxel during the block periods were extracted from the data. The block start and end times were shifted forward 4 seconds, to take into account the effect of the delayed haemodynamic response. To minimise the variability caused by the start and end of each block, the first and last volumes from each block were removed. For each voxel, the mean BOLD signal for each of the four directions was found by averaging the values from all the blocks of each direction over all the runs. This gave four values for each voxel, one for each of the four stimulus directions. This and the other analyses were performed using custom MATLAB scripts.

Partial Correlation

The main difference between the fMRI values from different voxels was the constant 'background' value of the voxel, not the variation induced by the BOLD signal. Failure to control for the background value would have led to the correlation between pairs of directions approximately reaching one, as the pair of direction values from one voxel will be approximately equal, when compared to the pair of directions from another voxel.

To control for this background effect, the partial correlation was used. To find the partial correlation between (for example) the response to upward moving dots and downward moving dots, one first needs to control for the left- and rightward responses. This was done by first regressing the upward and

downward responses, across all voxels, with regressors made up of the left- and rightward responses. The Pearson correlation of the two residual vectors (for the upward and downward responses) was then calculated in the normal way. This gave a correlation coefficient largely unaffected by the 'background' variation between voxels.

The difference in correlation between regions was compared using Bayesian Model Comparison using the method described by Wetzels et al. (2009). In our experiment, the null hypothesis was that the partial correlation between opposing directions equalled the partial correlation between orthogonal directions. The alternative hypothesis was that the correlation between opposing directions was greater. The method developed by Wetzels et al. (2009) allowed us to compare such order restricted hypotheses to quantify evidence for the null hypothesis. The Bayes Factor was calculated using the recommended parameters, suggested by Wetzels et al. (2009). We used a Cauchy prior, with two groups (of unequal variance).

2.2.5 Direction Preference Map

Another analysis investigated the existence of large-scale anisotropic maps. To determine if there were large-scale anisotropic direction-response maps in V1 and hMT+, the set of voxels most active for particular directions was calculated and rendered onto a flat-map of the cortex. These flat maps were then combined across participants to provide average across-subject response maps.

To calculate a voxel's normalised BOLD response for each direction, we used the voxel's mean BOLD responses to the four directions, z-scored and high-pass filtered. For each direction, those voxels with a normalised response more than 0.15 standard deviations above the mean were included in the binary response map for that direction (similar to the method described by Parkkonen et al., 2008). This resulted in four binary maps for each participant hemisphere, each showing which voxels had a strong response to that particular direction.

Freesurfer 4 (surfer.nmr.mgh.harvard.edu) was used to process individual subject anatomical images using the default recon-all preprocessing pipeline. This produced an inflated spherical map of the cortex for each subject. To allow the subject data to be combined, the individual subject data was registered onto the fsaverage surface using the Freesurfer tool mris_preproc (version 1.37).

Combining Images

For each direction (and hemisphere), the 30 subject binary response maps were averaged to create a new mean response map for each direction-of-motion, each cortical location having a value between zero and one indicating the proportion of subjects to have a strong response for that direction, at that location. The ROIs across subjects were also combined in the same way, giving a mean ROI map, in which the value at each cortical location indicated the ratio of subjects found to have the ROI at that location. Finally, the mean response maps were divided by the mean ROI map, giving a map illustrating the response strength in each direction, taking into account the variation in ROI shape across subjects. Any location that was outside more than 75% of ROI masks was ignored and left blank. This method

produced four images for each hemisphere, showing which parts of the cortex respond most to each stimulus direction (see Figure 2.9).

Statistical Analysis of V1 anisotropic map

Given the known anisotropy (Raemaekers et al., 2009) and retinotopic map (Engel et al., 1997) in V1, we hypothesised that the portion within the sulcus would have a greater response to horizontal motion, while the gyral portion would have a greater response for vertical motion. To confirm our detected anisotropy was significant, the response maps were divided into two parts, using the freesurfer sulco-gyral parcellation results from the recon-all pipeline. For each subject, the mean response from all V1 locations within the calcarine sulcus was calculated, as was the response from the portion outside the sulcus. These means were then compared across subjects.

2.2.6 Anisotropic Response

To test whether a bias exists in hMT+ we computed each voxel's preference to each direction of motion. These preferences were computed in precisely the same way as for the direction preference map. Each voxel's mean z-scored BOLD signal in the four directions was subtracted from the mean for each direction. The mean preference for each direction, averaged across all voxels (and all participants) was then calculated. This value, referred here to as the direction response was compared between directions using ANOVA and t-tests.

Analysis of Kaskan's Optical Imaging Study

The optical imaging study (Kaskan et al., 2010) provides eight images, one for each 45° direction (figure 2.3). It appears that nearby columns often represent opposing directions. To test this consider the correlation between a set of paired points a specific distance apart, we hypothesised that there would be different patterns of correlation depending on the orientations of the two directions being correlated and the distance between points. For example, nearby pixels will be very strongly correlated for the same direction, due to the columnar organisation of the cortex. At distances of a few hundred microns it seems opposite directions would be more correlated, as the two pixels of each pair will frequently lie in neighbouring columns. This fits extremely well with our hypothesis explaining the fMRI results.

To test this in a quantitative fashion, I found the correlation coefficient across pairs of pixels, for pairings of different distances. The pair of pixels was taken from images either of opposing motion directions or orthogonal. The correlation coefficient was calculated for a range of distances to reveal the pattern of spatial correlation.

2.3 Results

2.3.1 Evidence for directional information

A difference in correlation between regions could be confounded by the absence of direction-sensitive information in the BOLD response in one of the regions. However, Kamitani & Tong (2006) found that ensemble activity patterns in both hMT+ and V1 provide robust direction-selective information. To confirm these findings, multivoxel pattern classification with leave-one-out cross validation was used to classify the direction of motion. Both hMT+ and V1 had highly significant decoding results (across participants $p < 0.001$, see Figure 2.5D), indicating that both areas contained information about the direction of dot-motion.

2.3.2 Anisotropic Response

fMRI The mean direction response across the voxels of all participants for the four directions is plotted in figure 2.6A. It clearly shows that there are significant differences between the four directions (ANOVA, f=27.84, $p = 1.23 \times 10^{-13}$, tested across participants). With the two vertical directions having significantly greater average responses than the horizontal directions (see section 2.2.6 for details on the derivation of the direction responses).

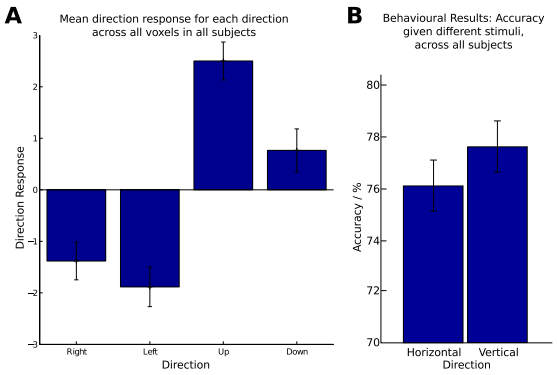


Figure 2.6: A. Mean direction response for each stimulus across all voxels in all participants. The error-bars represent one standard error (across participants). B. Behavioural accuracy in determining motion speed in the vertical and horizontal conditions, across all participants.

Once multiple comparisons have been accounted for (Tukey's test), rightward and leftward z-values are significantly different from upward or downward z-values (and upward are just significantly different from downward p=0.0037, t=-3.1556, corrected: p=0.02).

Combining the results into horizontal and vertical results, a t-test finds that vertical motion was represented more than horizontal (with t=-7.1045, 59 dof, $p = 1.8 \times 10^{-9}$).

Behaviour To shed light on these fMRI results it was useful to look at participant behavioural results. To assess whether vertical or horizontal motion speeds are more accurately distinguished, the proportion of correct trials in the four directions was found for each participant. Although the experiment was not designed to test for such anisotropies in perception, the vertical motion was correctly distinguished significantly more frequently than horizontal motion (figure 2.6B, one-tailed t-test across participants; p=0.032; t=-1.894).

2.3.3 Kaskan Optical Imaging Analysis

As discussed earlier, it seemed that the optical imaging results from Kaskan et al. (2010) (figure 2.3) show columns neighbouring those of the opposing direction. To test this, the correlation between different pairs of directions over different distances was computed.

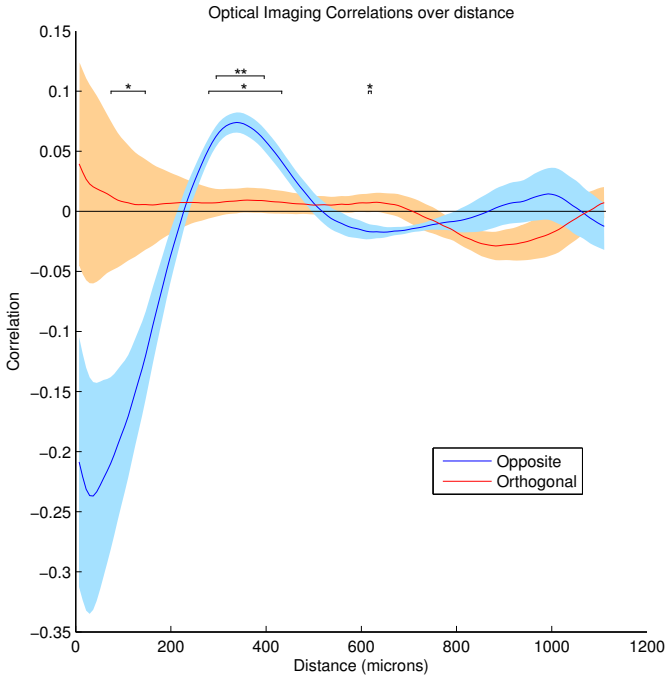


Figure 2.7: Analysis using data from Kaskan et al. (2010), figure 3. The correlation between pixels at different distances are computed. Blue, correlation between responses to opposing directions of motion. Red, correlations between responses to orthogonal directions of motion. Those distances in which the two correlations are significantly different (two-tailed, two-sample t-tests) are marked with asterisks. The coloured areas indicate one standard error confidence intervals, while the solid lines are the averages. These are computed across the available direction combinations (for example there are four possible pairs of opposing directions).

Figure 2.7 illustrates how the correlation between directions varies over distance. It shows that, at small distances (less than 200μm, i.e. within a column), opposing directions are negatively correlated.

At larger distances (between 300 and 450μm, i.e. between neighbouring columns), opposing directions are positively correlated, while orthogonal directions appear to be less correlated, on average.

2.3.4 Correlations

We hypothesised that hMT+ contains direction columns and that their organisation causes voxels to have preferences for opposing directions. To test this hypothesis, we correlated the responses to different directions of dot motion. Each voxel had four values, representing the mean BOLD responses to motion in the four directions. We calculated the partial correlation between pairs of these values, across all the 160 voxels in the ROI.

In hMT+ the average partial correlation between opposing directions (r = 0.4371, Pearson correlation) was far greater than between orthogonal directions (r = 0.1283, Pearson correlation, Figure 2.8A). In other words, if a voxel's BOLD response was greater for one direction of motion, it was more likely to be greater for motion in the opposite direction than in an orthogonal one. The average of the two opposing direction correlations was compared to the average of the four orthogonal correlations, across subjects (Figure 2.8B). A two-way, repeated measures Analysis of Variance (ANOVA) with the region (hMT+ or V1) as one factor, and the pairing of directions as the other (opposing or orthogonal) found a significant difference between regions (F(1,29) = 12.03, $p < 0.002$). More importantly, there was a highly significant difference between orthogonal and opposing pairs (F(1, 29) = 16.96, $p < 0.001$) with a large interaction effect between the two factors (F(1,29) = 29.74, $p < 0.001$).

In hMT+, a one-tailed paired t-test for the difference in these correlations found the two values significantly different when tested across participants ($p < 0.001$, t = 5.90. df = 29, upper panel, Figure 2.8B). The one-tailed paired t-test for the data from V1, between opposing and orthogonal motion found no significant difference (p = 0.11, t = 1.27, df = 29, lower panel, Figure 2.8B).

Under a frequentist hypothesis testing framework, one can only reject the null hypothesis, never the alternative hypothesis. So to determine whether or not the effect was present in V1 we used the Bayesian model comparison method described by Wetzels et al. (2009). The Bayes Factor between the null and alternative hypotheses was 1.87. This means that the null hypothesis (that opposing and orthogonal correlations in V1 are equal) was more likely than our alternative hypothesis, but not significantly (Jeffreys, 1961).

Could the hMT+ localiser preferentially select those voxels with preferences for opposing directions, confounding our result? While fMRI spatial smoothing and large receptive fields make this unlikely, we addressed this concern by using the first experimental session to localise hMT+ instead. This session exposed both visual hemifields to all four directions of motion equally, and therefore would not bias voxel selection. The new ROI was then used to analyse the remaining sessions. The significance of the correlation results was unaffected using the new ROI, therefore we assume the standard localiser did not confound voxel selection.

2.3.5 Large-scale direction response maps

The previous analysis established a correlation between opposing directions that was specific to area hMT+, which could be driven either by columnar organisation or by a large-scale anisotropic direction map. An anisotropic map would mean a greater response to some stimulus directions than others, and, as the cortex has retinotopic organisation, this may lead to some parts of the cortex having a greater response for motion in certain directions. This inhomogeneity would allow very small-scale preferences for opposing directions (e.g. due to bidirectional cells) to cause the correlation for opposing directions found in the fMRI BOLD responses. To investigate whether such patterns of anisotropy exist, we examined maps of responses to the four directions.

Images of the response maps were formed by combining data from each of the subjects. Each image (Figure 2.9) shows which regions of the cortex have the greatest response to motion in a particular direction. V1 was the region in which large-scale organisation was found and is discussed first.

As an example, we consider horizontal versus vertical motion. The majority of those voxels in V1 which have a greater response for horizontal (leftward and rightward) motion were found in the fundus of the calcarine sulcus (top four panels, Figure 2.9). The voxels with a greater response for vertical motion (upward and downward) showed the opposite pattern, with most of the increase in activity restricted to the upper banks of the sulcus (lower four panels, Figure 2.9). To compare these two regions (defined anatomically, see Materials and Methods), the proportion of each region highly responding (over 0.15 standard deviations above the mean) to each stimulus was calculated. Across all subjects, sulcal V1 had 28% more cortex highly responding than the gyral V1 when motion was horizontal, while gyral V1 had 32% more highly responding cortex than the sulcal V1 when the motion was vertical. This was tested using a repeated measures 2-way ANOVA, separately for each hemisphere and was found to be highly significant (interaction effect of sulco-gyral region vs stimulus direction: left hemispheres, F=8.67, $p < 0.001$; right hemispheres, F=7.19, $p < 0.001$).

In contrast to V1, a response map generated using the same methods using data from area hMT+ did not reveal any anisotropy in the response patterns (see Figure 2.10). However, for several reasons we cannot conclude that this implies that a large-scale map is absent. First, hMT+ is far smaller than V1; second, it is in a very folded and convoluted portion of the cortex making segmentation much more difficult; third, its location and orientation are much less stable across subjects, making alignment much less likely. It is the results of the study by Raemaekers et al. (2009) which found an absence of any anisotropy which motivates our assumption about the lack of such a bias in hMT+. See section 2.4.2 for details.

2.4 Discussion

In studies in non-human primates, neurons with preferences to opposing directions of motion were found to be organized in adjacent columns (Albright, 1984; Malonek et al., 1994). These findings led us to hypothesise that voxels in regions which contain direction preference columns would have a greater

BOLD response for opposing directions, than for orthogonal directions. Consistent with this hypothesis, we found that voxels in hMT+ had more correlated responses for opposing pairs of directions than for pairs of orthogonal directions, an effect which was absent in V1. Alternative explanations, discussed below, require an anisotropic map. We looked for such maps in both V1 and hMT+ and found one in V1, the region in which the correlation effect was absent, while no map was found in hMT+, where a robust correlation effect was found. This double dissociation indicates that the correlation effect is not primarily caused by a large-scale map organisation. Rather, we suggest that the preference for opposing directions is an indicator of columnar organisation.

2.4.1 Alternative explanations for the correlation between opposing directions

One might initially conclude the correlation for opposing directions is caused by bi-directional cells, a subpopulation first reported by Albright (1984), in which each cell has a preference for two opposing directions. Bidirectional cells have been found primarily in region MT, and are reported to make up 7 or 8 percent of the neurons in MT (Felleman & Kaas, 1984), in owl monkey and macaque. The two peaks of the cell's firing rate response curve are reported to be generally equal. Alternatively, one might imagine the effect could be due to nearby opposing inhibitory interneurons (Snowden et al., 1991), especially as some research shows hMT+ has greater motion opponency than V1 (Heeger et al., 1999). These interneurons are paired with others of the opposing direction of motion and so activity in this local area of cortex would be greatest when the stimulus direction is in either one of two opposite directions. Both these explanations (bi-directional cells and interneurons) describe a small-scale correlation between opposing directions. They don't, however, explain how correlations between opposing directions can be detected at the scale of voxels, where such small-scale inhomogeneity would remain undetected. Such small-scale preferences for opposing directions must be combined with a larger-scale inhomogeneity to cause the correlation effect observed at the scale of fMRI voxels.

These inhomogeneities may be present at any of three scales. The first, at sub-voxel scales, is our columnar-organisation hypothesis, which has already been discussed. The second, at the scale of several voxels, consists of larger 'blobs' of cortex having a preference for one particular direction over the others. The third scale is a large-scale, anisotropic retinotopic organisation, in which cortical areas that contain neurons with receptive fields in one part of the visual field will have a bias for motion in a certain direction. For example, a centrifugal (outward) motion anisotropy would cause voxels with receptive fields at the top of the visual field to have a bias for upward motion. I originally argued that without such an anisotropy, it is unlikely that small scale effects of bidirectional cells or similar could cause voxel-scale correlations between opposing directions of motion. However, I realised that by including MST in the ROI there is another alternative. Although not at the scale of columns or maps, the clustering of cells in MST by type has been reported. Tanaka & Saito (1989) described three types of cell in MST, *Direction* (D) cells, *Expansion/Contraction* (S, for 'size-change') cells and *Rotation* (R) cells. These results were confirmed by Duffy & Wurtz (1991) who found cells often preferred combinations of these stimuli. Importantly for this discussion they found that neurons of the same class often were in

close proximity. The clustering of 'size-change' cells could therefore be an alternative explanation for the correlation results, and does not depend on traditional columns. Further work is needed, possibly through both modelling the predicted results of the columnar organisation and carefully segmenting the hMT+ region to refine this technique and prove that the effect is due to columns and not alternative structures.

2.4.2 Review of large scale direction anisotropies

We looked for large-scale anisotropic maps in V1 and hMT+, which might explain, due to the mechanism described above, the correlation between opposing directions. If these are the explanation, one would expect the correlation to be strongest in those regions with the most robust directional anisotropy. For example, (Freeman et al., 2011) conclude that decoding orientation from V1 depends on an anisotropy across V1. They suggested that this is caused by retinotopic organisation, combined with a centripetal or centrifugal anisotropy in orientation or direction. Therefore, our search in V1 was guided by the retinotopic map described by Engel et al. (1997) and the anisotropy discovered by Raemaekers et al. (2009). Evidence for an anisotropy in hMT+ is far more equivocal. The same study (Raemaekers et al., 2009) found no anisotropy in hMT+, contradicting an earlier study (Giaschi et al., 2007) which did. It was suggested by Raemaekers that Giaschi's results may be due to the use of near-threshold motion, which would induce attentional effects to motion. Our experiment used 100% coherent dot motion, and so we would expect our results to match those of Raemaekers and no anisotropy to be present for our stimuli in hMT+. Our results reveal the predicted anisotropic direction map in V1, which fits with the anisotropic bias described by Raemaekers et al. (2009) and the retinotopic map detailed by Engel et al. (1997). For example, combining the results of these two studies would lead to the prediction that the fundus of the calcarine sulcus would have the greatest response to horizontal motion. This is precisely the result found (see Figure 2.9, top panels). In contrast, no organised anisotropic preference map was apparent in hMT+. In comparison to V1, hMT+ is not as well defined anatomically. It is smaller and is found in a region of heavily folded cortex. It is possible the absence of a detected organised map is due to these difficulties analysing this region, but it does match Raemaekers et al. (2009) finding that there is no anisotropy in hMT+.

The double-dissociation described above between anisotropic direction maps and the correlation effect, appears to confirm our hypothesis that hMT+ has columnar organisation, but could an undetected large-scale direction map be responsible for the correlation effect? A coarse retinotopy has been described in hMT+ and MT (Huk et al., 2002; Wandell et al., 2005, 2007; Xu et al., 2004), but to generate the inhomogeneity one also requires an anisotropy which we have already dismissed above. Further evidence against a small-scale organisation for opposing directions is Xu et al. (2004)'s observation that 'movement of an oriented stimulus in one direction did not activate the same territories [in primate MT] as movement in the opposite direction'. This is further evidence which strongly suggests that the small-scale positive correlations between opposing directions, required by this alternative explanation, do not exist.

2.4.3 Review of columnar organisation from animal studies

It is likely that humans have direction columns in hMT+, given that other primates have been found to have such a map (Xu et al., 2004; Malonek et al., 1994; Kaskan et al., 2010; DeAngelis & Newsome, 1999). Whether humans have a similar columnar direction-preference map in V1 is still disputed. Optical imaging has found columnar direction preference maps in some animals, such as cat (Shmuel & Grinvald, 1996), ferret (Weliky et al., 1996) and marmoset (Roe et al., 2005), while in macaque none were found (Lu et al., 2010). As discussed by Lu et al. (2010), electrophysiological recordings found frequent reversals in direction preference during vertical penetrations of macaque V1 (Hubel & Wiesel, 1968), suggesting an absence of an organised directional map. Macaque are the most recent common ancestors of humans to have had V1's direction responses investigated in such a way. It is quite plausible therefore that there is no direction columnar organisation in human V1. In non-human animal studies investigating MT, columns of opposing direction preference are often found next to each other. For example, electrophysiological recordings by Albright et al. (1984) in macaque, found many 180° changes in direction preference during tangential penetrations, suggesting neighbouring columns often prefer opposing directions. Malonek et al. (1994) also found that direction columns appeared to neighbour columns preferring opposing directions using optical imaging from owl monkey. They noted that orientation patches appeared to be divided into two, with each half preferring motion in opposite directions. Our analysis of the optical imaging by Kaskan et al. (2010) also supports these conclusions.

2.4.4 Motion streaks

Finally, there is the question of whether the correlation of opposing directions could be due to motion-streaks. Motion streaks are induced when dots translate quickly across the visual field, during fixation (Geisler et al., 1999). If this were the explanation, the correlation would be due to a region or column with a particular orientation preference, responding similarly to motion in either opposing direction. V1 neurons are reported to have strong orientation selectivity (Hubel & Wiesel, 1968), while hMT+ neurons are thought to be far more selective to the direction of motion (Albright, 1984). It seems unlikely that motion streaks are the explanation, as the correlated opposing directions effect was absent in V1 exactly where one would expect it to be strongest, if motion streaks were the cause.

In summary, although alternative explanations exist for the correlation of opposing directions, the proposal that opposing columns explain the result is highly compelling. With a little more research it will probably be possible to show that the correlation of opposing directions in functional imaging data is a sign of columnar organisation. It should be noted that this model depends on an organisation in which columns with opposing directions are adjacent. Columnar organisation would remain undetected by our method, if columns were organised in a different fashion.

2.4.5 Preference Bias

Aside from the main findings of the study, we also found that there were significant anisotropies in both the fMRI and behavioural responses, with vertical directions inducing greater BOLD signals than horizontal. The behavioural results show that vertical motion speed discrimination was significantly easier for participants than horizontal.

At first we considered whether such a bias could be due to the equipment or stimulus; maybe the screen was not angled correctly, thus causing motion in one direction to appear to be faster than in another. However the angles of the mirror screen and projector were checked and confirmed to be correctly orientated.

The results appear to fit the hypothesis proposed; that the increased prevalence of vertical motion has caused the cells in hMT+ to have a preference for vertical over horizontal motion. This increased number of cells will be detected by an increase in the BOLD signal in the vertical direction. This hypothesis was based on that of Hansen & Essock (2006) and Girshick et al. (2011), who proposed that differences in perceptual accuracy are caused by a variations in the population of cells tuned to those stimuli. However, there are other possible explanations for these results. An intriguing study found that vertical motion was perceived to be quicker than horizontal (Scott-Samuel & Magapu, 2002). The possible weakness in distinguishing the direction of vertical motion (compared to horizontal) hinted at by the figure in Gros et al. (1998), could be explained by this anisotropy; if the motion in the horizontal direction is perceived to be less, then the angular difference in the vertical direction will be reduced, compared to the same (real) angle in a horizontal direction. Thus the difference in the ability to distinguish direction in the two orientations could be driven by the same process. It is known that higher speeds are associated with an increase in the fMRI BOLD signal in region hMT+, thus it might be that the increase in the *perceived* speed is either explained by or related to the increase in the BOLD signal from the region. Finally, our own behavioural results show that vertical speed differences are more easily distinguished than horizontal. This would also fit with a greater proportion of cortex sensitive to vertical motion over horizontal motion.

The slight fMRI signal preference for upward over downward motion needs an explanation. It is intriguing that a bias exists between upward and downward motion and not leftward/rightward. This appears to match the symmetry in the horizontal direction and asymmetry in the vertical, in natural scenes. If we assume that the relative strength of the BOLD response is linked to the prevalence of the stimulus in natural scenes, then these results lead to the hypothesis that upward motion is more common than downward in natural scenes, for instance during optic flow while walking. Possibly conflicting with this interpretation is the finding from an electrophysiological study in macaque (Maunsell & van Essen, 1987) which found that the lower visual field (presumably where the ground is usually found) is over represented in the retinotopic map in MT. Clearly this area requires additional research to resolve this apparent contradiction.

In conclusion, the vertical/horizontal difference in fMRI representation and psychophysical ability appears highly significant, and, in general, appears to correspond to the hypothesis of a greater repre-

sentation of vertical motion in region hMT+. Other plausible explanations may exist for these results. Future research however should concentrate on oblique angles as one can probably resolve the same questions with greater signal strength using these orientations.

2.4.6 Kaskan Optical Imaging

The optical imaging data from Kaskan was found to have a negative correlation between images of opposing motion directions, at distances of about $400\mu m$. Although the scale of these correlations appear to fit our fMRI correlation results, it is still a qualitative observation. These results could be extended considerably by modelling the correlation one might expect if fMRI voxels were placed over the optical imaging preference map. Much larger data sets are required to provide the model its raw data. Alternatively one could use simulated preference maps (e.g. that produced by Topographica[1]). This author found, however, that the maps produced didn't have the neighbouring opposing direction column organisation, possibly because it was modelling V1 and not hMT+.

Once a large enough source data set has been obtained, the fMRI voxels could be simulated at different scales, by using voxels of different sizes when averaging the preference maps. The changes in correlation over these different scales could then be compared to the change in correlation over different scales for the real fMRI data. By using a 7T fMRI scan one could obtain a greater range of voxel sizes (e.g. 0.5mm to 5mm). By fitting the modelled and actual variations in correlation it should be possible to estimate the size of columns in human MT+.

2.4.7 Future Research Directions

An important question that needs answering is exactly how columnar organisation produces the opposing correlation reported. Analysis of maps from optical imaging studies may help reveal which features of the columnar organisation are responsible. A potential method as mentioned would be to simulate the sampling during fMRI using data generated by optical imaging. This may allow us to determine features such as the scale of the columns, from the fMRI data.

The hMT+ complex has been found to include two main subfields, the human putative middle temporal (hMT) area and the human medial superior temporal (hMST) area. It is likely to include other motion-sensitive regions too (Kolster et al., 2010). Although our hypothesis was based largely on the known organisation of non-human primate MT, the experiment did not distinguish the subfields of hMT+, and so the results are currently not specific to one of the subregions. For example the columnar-organisation causing the correlation may be found in hMST rather than hMT, which has also been reported to have evidence of columnar direction organisation in non-human primate electrophysiological recordings (Britten, 1998). Distinguishing the subfields responsible for the correlation should be targeted in future studies. Several methods have been developed to distinguish these areas (Huk et al., 2002; Smith et al., 2006, e.g.). If the correlation effect varies between these subfields, it could be used to test which area a voxel is in. This could complement the current methods and may improve the precision

[1] http://topographica.org/Home/index.html

with which the region is segmented. This paper's results regarding the correlation between opposing directions may shed light on a recent debate in the field of multivoxel pattern analysis (MVPA). It was widely assumed that the variation in voxel BOLD responses was due to columnar organisation (Kamitani & Tong, 2006). It was supposed that even with randomly distributed columns, a voxel is likely to contain more columns preferring one direction. This would lead to an overall preference in that voxel's BOLD response for motion in one direction. However, Freeman et al. (2011) found large scale retinotopic orientation maps across V1, potentially explaining the source of variation between voxels required by MVPA, without invoking columnar organisation. Our results suggest that columns do influence the response of individual voxels. Therefore, it is likely that MVPA can be assisted by the presence of columnar organisation, but retinotopic maps can also assist decoding. To summarise, both maps and columns are important for MVPA, depending on the region in question.

2.4.8 Implications for self motion processing

The results suggest that some part of the human hMT+ complex has columnar organisation, structured in such a way that neighbouring columns respond most to opposing directions of motion. This pairing may aid the development of inhibitory connections between the pairs of directions. Such paired inhibition is commonly cited as an explanation for the waterfall effect and other adaptation-related phenomena. Investigating and using adaptation as a tool (especially in less investigated domains, such as auditory motion stimuli) may provide interesting results. The following chapter contains a behavioural adaptation element, with both auditory and visual self-motion being simulated and the adaptation effect investigated through the number of erroneous responses.

The finding of an anisotropic map in V1 fits well the retinotopic organisation of the region and the reported psychophysics anisotropies. It is likely that the columns for the direction of highest response will be larger (rather than more densely populated or have higher firing rates), based on the evidence from optical imaging (e.g. Chapman & Bonhoeffer, 1998). Such anisotropies are likely to exist in many domains. The hypothesis that these are a reflection of the statistics in the subject's developmental environment can be tested by looking for more examples of such anisotropies, for example in auditory motion and possibly in the field of integration. Priors, such as the preference for visual stimuli over auditory, could be cited as an example of such environmentally learnt anisotropies. The study by Putzar et al. (2007) found that those who were blind during their first few months of their lives showed much reduced visual capture effects in later life, even though they shared similar visual acuity to controls during the experiment. The experiment in the following chapter will briefly investigate visual capture through a final 'conflict' trial in which the two stimuli disagree.

The anisotropic map found in V1 suggests the region responds most to inward or outward motion; stimuli strongly associated with selfmotion. It seems that even at the level of V1 the effect of self-motion on development is important. It would be interesting to repeat the anisotropic analysis but with the cataract patient group from the study by Putzar et al. (2007), to test whether the anisotropy is caused by some feature in early visual stimuli. Although it seems to this author that much of a human's early

visual experience is not forward self-motion, maybe the anisotropy developed through expose to retinal-waves instead, or through later stimuli once self-motion was available.

The experiments in the next two chapters extend the stimulus considerably by introducing more realistic self-motion and incorporating an auditory stimuli. The focus shifts from the structure of unimodal brain regions to the connection between those which may be involved in integration. Models such as Bayesian integration are tested both behaviourally and using fMRI. In particular, the Normalisation model (by Ohshiro et al. (2011)), a suggested organisation of neurons which could underpin Bayesian integration, was modified to allow it to predict the psychophysics results. Paired inhibition was not included in the original model or in the implementation in this thesis. However, more details, using experiments like those in this chapter could allow estimates to be made about such local connectivity. Inter-subject variability, in particular in aging or disease states might provide considerable clues regarding the changes in the organisation.

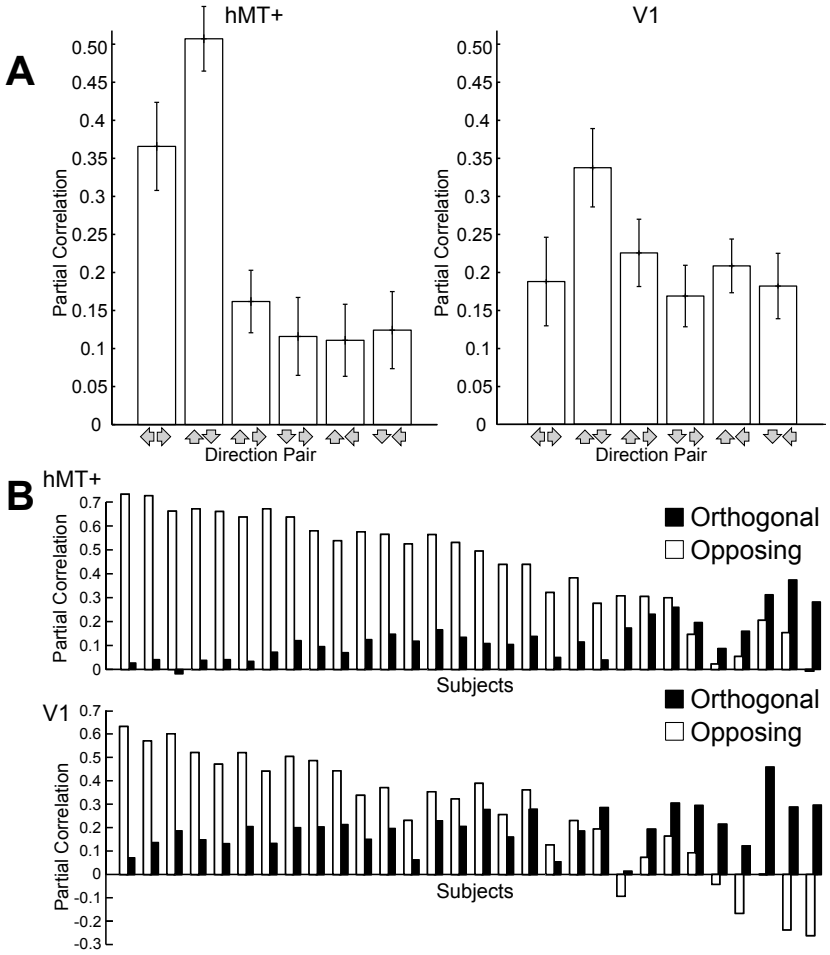


Figure 2.8: A. Average of the partial correlations of voxel BOLD responses across all subjects for each of the six possible pairs of stimuli directions, for hMT+ (left) and V1 (right). Error bars show the standard error across subjects. B. Average partial correlation coefficient for each subject, for those pairs of directions which are opposite (white) and those pairs of directions which are orthogonal (black). The subjects have been sorted separately for each graph by the difference between the opposing direction and the orthogonal direction. The graphs show the results for the two regions: hMT+ (top) and V1 (bottom).

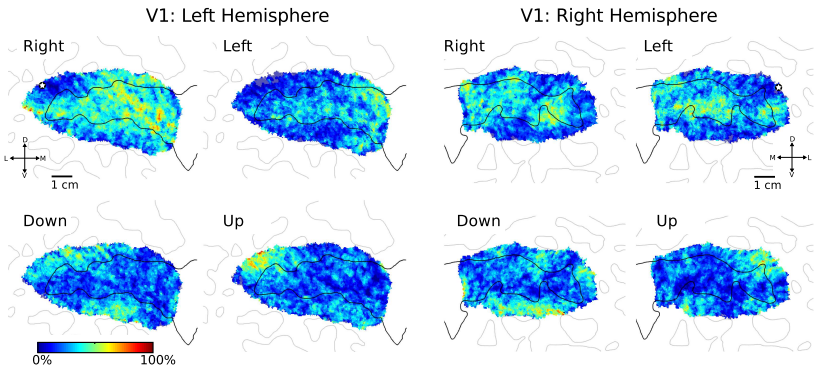


Figure 2.9: V1 direction preference maps created by combining masks from each subject. The maps show the distribution of responses for different directions of motion. The left four figures are for the left hemisphere, the right four are for the right hemisphere. Each of the four images shows the responses to a different stimulus direction, clockwise from top left: rightward, leftward, upward and downward. Blue indicates that only a small proportion of the subjects had high activity (greater than 0.15 standard deviations) at a location while red indicates that many subjects did. The black line marks the edges of the calcarine sulcus for the standardised fsaverage subject (grey lines indicate the boundaries of other sulci). The compass indicates the directions: D, Dorsal; V, Ventral; L, Lateral; M, Medial. The scale bars are 1cm long. The star indicates the approximate location of the occipital pole for the standardised fsaverage subject.

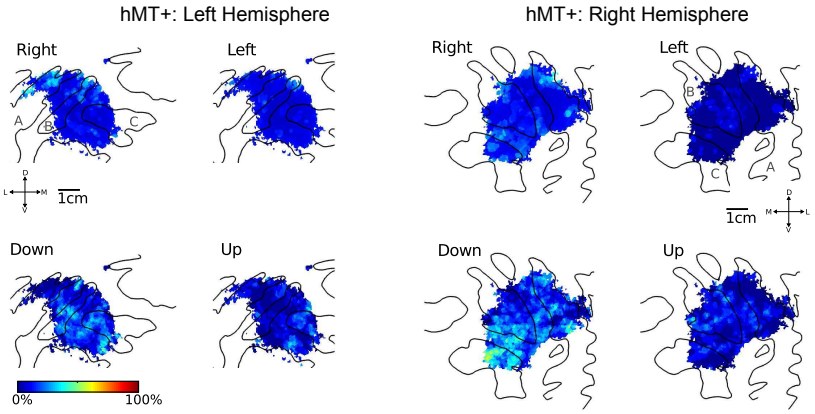


Figure 2.10: hMT+ response maps created by combining masks from each subject. The compass indicates the directions: D, Dorsal; V, Ventral; L, Lateral; M, Medial. The scale bars are 1cm long. As with Figure 4, the maps show the distribution of responses for different directions of motion, using the same colour scale. Unlike the V1 maps there is no obvious anisotropy. The three labelled sulci are A, the superior temporal sulcus; B, the ascending limb of the inferior temporal sulcus; C, posterior continuation of the inferior temporal sulcus.

Chapter 3

The Psychophysics of Audiovisual Selfmotion Integration

The planar-motion stimulus in the last chapter is only part of the optic-flow cue, and as such represents only a fragment of the stimuli associated with self-motion. In this chapter the integration of more complex self-motion cues is investigated in a psychophysics experiment.

The experiment devised was also intended to prepare the fMRI study to follow. In the previous chapter the whole hMT+ complex was included in the analysis, although we were most interested in the human equivalent of the primate MT region. Currently, studies suggests that MT has no preference for selfmotion congruent stimuli and can be considered to be principally receptive to small patches of planar optic flow (Wall & Smith, 2008). MT outputs to MST (a neighbouring region) (Britten, 2008), which has a greater response to self-motion cues (e.g. expansion). We hypotheses that this self-motion cue signal is integrated with others (including auditory motion cues) in VIP, a region investigated later in chapter 4. The output of this region and other integration regions is then involved in the modulation of the head-direction system found in the hippocampal formation, investigated in chapter 5. In this chapter we focus on the perception and behaviour surrounding this integration. In particular, we attempt to show whether the integration of the two cues is bayes-optimal, or if not, which alternative model explains the pattern of results most accurately. We also investigate adaptation - the influence of previous trials on the results of the current trial. Although not an initial design feature, the adaptation effect appeared to play a very strong role in the results. A rudimentary analysis was performed to look at whether the adaptation results give clues about the process underpinning integration.

3.1 Introduction

Many activities critical to an animal's survival involve fast and precise self motion. To achieve this, separate sensory inputs which contain self-motion information need to be accurately integrated. One example of integrating selfmotion cues is the ambiguity in the signal provided by the otolith organs, the vestibular system's linear accelerometers. The equivalence principle means that these organs are unable to distinguish between some orientations and accelerations. We clearly can not depend on these organs alone when trying to remain upright or walk. Other senses, including visual, tactile and proprioceptive feedback are required to resolve these ambiguities. For example, optic-flow is directly used in maintaining balance and invalid optic flow cues can cause young children to fall over (Lee & Aronson, 1974). In general, even when a cue can provide an unambiguous estimate of a environmental feature, combining its estimate with others will further reduce uncertainty.

Navigation involves keeping track of several, related, directions. Briefly, the *course* a person is on is the direction they are travelling, relative to some allocentric reference frame, along which they travel at a certain *speed*. Their facing direction or *heading* doesn't necessarily need to be in the direction of their course (for example if they turn their head to look at a tree besides the path) (Loomis et al., 1999). Computations comparing and subtracting angles from one another may involve a dedicated series of neural systems (Zacks, 2008) in or near the intraparietal sulcus (among other areas). The focus in chapter 4 on the ventral intraparietal sulcus reflects the apparent importance of the area in spatial processing. A basic step in any navigation problem is the ability to determine the difference between one's heading and one's bearing. Several cues might be used to calculate this angle, but one of the most salient may be the optic flow cue. As mentioned in the previous chapter, optic flow is the projected pattern of motion onto the retina as an organism moves (Gibson, 1950). The calculation of one's trajectory using the location of the centre of expansion and surrounding motion stimuli is not completely trivial however, as the trajectory of one's gaze must be taken into account. This is known as the rotation problem. This problem is investigated extensively elsewhere and isn't looked at explicitly in the following two chapters. During the fMRI scanning the participants are requested to fixate on a central fixation cross, minimising any extra cognitive load solving the rotation problem might induce. Regardless of the rotation problem's particular implementation, a typical step in estimating the direction of motion (at least in the case with a fixed gaze direction) would be to estimate the direction of the centre of expansion. Intriguingly VIP in monkeys has been found to contain cells with very wide receptive fields that are particularly receptive to self-motion stimuli, with peak responses which vary depending on the direction of the centre of expansion (Gu et al., 2006; Zhang & Britten, 2010). The human homologue of this region appears to also be involved with self-motion related stimuli in humans (Wall & Smith, 2008). Although the rotation problem could be solved purely by incorporating the eye's rotation trajectory, humans seem to also use depth cue information (Van den Berg, 1992), in particular tracking points at infinity (i.e. the horizon) seems to be important during rotations. Using a plane rather than a dot 'cloud' was found to provide better estimates of heading direction; human estimates of heading direction, using such a stimulus is reported to be accurate to within two degrees (Warren et al., 1988). These speed and rotation

cues can be used to estimate one's trajectory, without recourse to allocentric cues, a useful skill in a poorly lit, occluded or featureless environment. Dead reckoning or *path integration* is used to update the current estimated location using the cues regarding self motion. Although essential, path integration accumulates errors so the estimates it provides must be supplemented with more accurate allocentric cues at least occasionally, to allow the current location to be more accurately identified. Integration like this seems to be required across a wide number of tasks, both within self-motion and beyond.

The following two chapters investigate the integration of visual and auditory selfmotion. Auditory motion cues have been found to evoke responses in VIP neurons in macaque (Schlack et al., 2005), and the region has been found to respond to stimuli of many modalities (e.g. Zeki et al., 1991; Colby et al., 1993; Bremmer et al., 2002b; Duhamel et al., 1998). It hasn't yet been shown that VIP responds preferentially to auditory *self motion* stimuli, but it is a reasonable working assumption, given the likely importance of self-motion to the region in other modalities.

It is important not to assume the integration is restricted to VIP (this hypothesis is tested in chapter 4). Other regions may provide similar functionality. For example MST, besides responding to complex visual stimuli, responds to vestibular cues (Fetsch et al., 2007) and pursuit eye-movements (Newsome et al., 1988). However, fMRI studies haven't found MST responding to auditory motion in (sighted) adults (Bedny et al., 2010; Lewis et al., 2000). In this chapter, the analysis and discussion is based on a simple idea of two unimodal regions (visual and auditory) and an integration region where these cues are combined. This model is considered and tested in more detail in chapter 4.

3.1.1 Background to Integration

A large body of literature now exists investigating how different sensory modalities are integrated. The integration of disparate stimuli seems to be a ubiquitous feature of human perception. Famous examples include the McGurk effect (a form of crossmodal interaction), in which a perceived audible syllable is altered by simultaneously seeing a person speak a different syllable (McGurk & MacDonald, 1976). Other examples include such pairings as visual and haptic integration (Ernst & Banks, 2002), sound and visual human-movement stimuli (Scheef et al., 2009) and the visual and auditory 'ventriloquist' effect (Alais & Burr, 2004).

Integration vs Interaction

The literature can be divided into two groups depending on the approach taken to combining cues. Spence et al. (2009) describe the two approaches as either crossmodal integration or multisensory integration. The former refers to those situations in which one stimulus influences the perception of another (for example though attention), while multisensory integration is the slightly more general term which looks at the response of either organisms or individual neurons to the combining of various cues (Spence et al., 2009). A quote from Schmiedchen et al. (2012) (citing Spence et al. 2009), explains the same division:

> multisensory integration refers to the binding of stimuli perceived through multiple senses, whereas crossmodal interactions describe the direct influence of one modality on signal processing in another modality without necessarily integrating information (Spence et al., 2009).
>
> *(Schmiedchen et al., 2012)*

Whether these are, in fact, the same process or consist of two separate mechanisms is unclear. If the latter, we must also ask if or when one takes over from the other in different circumstances. This experiment aims to concentrate on the integration of stimuli, rather than just the interaction of stimuli. Whether this is a possible or desirable experimental design are also open questions.

Meredith and Stein's rules for integration

A brief review of the history of the field can help explain this division. Initially, evidence from the ventriloquist effect (McGurk & MacDonald, 1976) suggested that visual-capture was the principle process, supported by several studies, for example after investigating the effect of auditory cues on visual apparent motion, Allen & Kolers (1981) wrote that 'events in one modality affect events in another' without a 'suprasensory' integrating layer being required. The field of electrophysiology was also having an influence. Meredith & Stein (1986), while investigating the activity of the superior colliculus in cats, developed three general principles of integration: the spatial rule (unisensory stimuli at same location), the temporal rule (unisensory stimuli at same time) and the rule of inverse effectiveness (multisensory integration is strongest when unisensory stimuli produce weak responses when presented individually).

Bayesian Integration

Since then, further research has caused opinion to shift. The paper by Ernst & Banks (2002) marked a seminal moment in the field, with most of the research in the last ten years performed through the prism of Bayesian integration. Briefly, Bayesian integration is a method for combining different estimates of a parameter, taking into account the uncertainty of the estimates in such a way as to pick the most likely value. It does this using Bayes' rule, a simple statement about probabilities, discovered by Thomas Bayes (see Box 1).

The use of Bayes' rule to explain how we weight the evidence from different stimuli or cues has been remarkably successful, explaining a wide variety of cue integration results (Pouget et al., 2002). It seems likely therefore that it should apply to the integration of the two stimuli used here.

Looking again at Visual Capture

The Bayesian explanation for visual capture is that the visual cues are more reliable than the auditory, and so will be weighted much more strongly than the auditory cue (Alais & Burr, 2004). If Bayesian integration is the explanation though, it should be possible for the converse situation to arise. If the visual cue's precision is degraded sufficiently, one should expect 'auditory capture'.

Box 1 Bayes' Rule example

Bayes' rule is most commonly expressed as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{3.1}$$

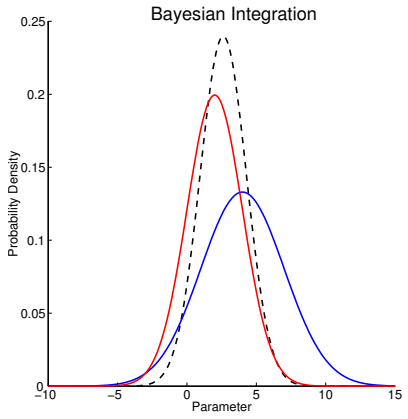
Where $P(A)$ and $P(B)$ are the probabilities of events A and B. $P(A|B)$ is the probability of event A given that event B has occurred, while $P(B|A)$ is the probability of event B given event A has occurred.

A simple example: A crime (stolen jam) has taken place in a small town of 1000 people. A suspect has been found with jam on their face. The probability of any individual having committed the crime is 1 in 1000. On average, the probability of any one person having jam on their face is 1 in 100, while the probability of a thief having jam on their face is one in four. Putting these values into the equation:

$$P(thief|jam) = \frac{P(jam|thief)P(thief)}{P(jam)} \tag{3.2}$$

$$P(thief|jam) = \frac{25\% \times 0.1\%}{1\%} = 2.5\% \tag{3.3}$$

So there is only a 2.5% chance the suspect stole the jam. In this example the probabilities were binary (true/false), while in the Bayesian integration considered in this chapter the values of the estimates are considered to be continuous variables. The above equation still applies, but the probabilities are calculated across ranges of values, for example $P(A = a|B = b)$.



The figure shows an example of Bayesian integration. The blue and red lines are different estimates of a continuous variable. The dashed black line is their product (assumes equal prior weight to each estimate). Note that this new curve is narrower than the two estimates, which means it has a smaller variance in its estimate.

So, can auditory stimuli ever influence perception more than visual stimuli (reversing the ventriloquist effect)? Hidaka et al. (2011) found that an auditory motion stimulus could cause the perception of motion in static visual stimuli, outside the fovea. The low visual resolution outside the central visual field combined with the reliable auditory stimulus may be key to this finding. Hidaka et al. (2011) found that eccentricities of over 10 degrees allowed such auditory capture. Our experiment is of this order (up to 20 degrees from the fixation cross), so this may help in encouraging Bayesian integration, without a bias for visual motion. More important for this study are the results from their second experiment, with a global field of dots. There they found that, '*contrary to the previous studies...the current results clearly demonstrate that continuous lateral shifts of sound can induce visual motion direction perception consistent with auditory movement in a global motion display.*' (Hidaka et al., 2011)

The Normalisation Model

A model developed by Ohshiro et al. (2011) offers a plausible neural system which produces near Bayesian integration. The model has a layer of neurons at each of its two inputs. Each layer, i, is unimodal. The responses, $I_i(x_j)$, of neurons in these layers increase sublinearly with increasing stimulus intensity, x_j, and it is assumed that neurons in the two regions have overlapping receptive fields. The two unimodal regions are connected to a multisensory integration region in such a way that neurons with the same receptive field in the two unimodal regions are combined. The multisensory neurons' inputs are a weighted sum of the unimodal outputs (equation 3.4).

$$E_j = w_1 I_1(x_j) + w_2 I_2(x_j) \tag{3.4}$$

The dominance weights w_1 and w_2 vary between neurons depending on the dominance of one modality. The output of the integration neuron, R, follows an expansive power-law (simulating the function linking membrane-potential to firing rate). Finally, the activity of each neuron is divided by the sum of the activity of all multisensory neurons to produce the final response (equation 3.5).

$$R = \frac{E^n}{\alpha^n + \left(\frac{1}{N}\right)\sum_{j=1}^{N} E_j^n} \tag{3.5}$$

Where parameters n and α need selecting. The result could be chosen to be the bimodal neuron with the highest firing rate.

This model appears to closely follow Bayesian integration results, while being built on plausible neural responses. The model can also explain such phenomena as multisensory enhancement and inverse effectiveness. To use it in our psychophysics study however requires modifications to generate reasonable input neuron activity. These changes are detailed in section 3.3.5. An alternative competing model, proposed by Ma et al. (2006) relies on the Poisson nature of neural activity and a linear combination of population codes to produce Bayesian integration-like results. The key differences between the models is related to superadditivity and how changes in reliability would affect the total firing rate within the unimodal regions. The Ma model is not tested in this study, but it is important to note that the Normalisation model used here is not the only plausible model of integration. The modifications and

assumptions made for the Normalisation model to adapt it for psychophysics analysis could equally be applied to the Ma model and others.

3.1.2 Self Motion cues and their integration

Studies have investigated the integration of visual and vestibular cues for self motion (e.g. DeAngelis & Angelaki, 2012). There have been a reasonable number of studies into audio-visual motion (e.g. Schmiedchen et al., 2012; Hidaka et al., 2011; Wuerger et al., 2003; Meyer & Wuerger, 2001) but none of these studies have used specifically self-motion cues.

Of particular importance to our study are reports that auditory stimuli have been found to induce sensation of selfmotion (an effect referred to as *auditory vection*). Attempts to induce circular vection go back to the 19th century (e.g. Urbantschitsch, 1897), while linear vection is far less researched (e.g. Soames & Raper, 1992; Väljamäe et al., 2008). The discovery of auditory vection suggests that auditory motion is a valid and important stimulus for self motion, and is probably included in the cortical integration of self-motion cues.

Several experiments have looked at self-motion integration. That is, the integration of cues to estimate aspects of the subject's physical motion (particularly with respect to forward directed movement). Examples include looking at how humans weight proprioception inputs and landmarks (Nardini et al., 2008), vestibular and visual cues (Fetsch et al., 2009). Other features related to self motion include visual and auditory integration of depth cues (Karaoguz et al., 2011). However, this author knows of no studies which investigate the integration of visual and auditory *self motion* cues. The most relevant is a brief study which found the attention to the direction of a visual motion cue influences the response time to an auditory motion cue (Beer & Röder, 2004), a promising hint that integration between the two stimuli will be possible. It should also be noted that the results of the studies mentioned previously, looking at non-self-motion audio-visual integration, may be applicable to self-motion cue integration. For example the results of Meyer & Wuerger (2001) indicate that visual and auditory horizontal motion is integrated while more recent papers (e.g. Alais & Burr, 2004) more precisely suggests the integration is Bayesian. This experiment was devised to test the inference that self motion stimuli would also be integrated in a Bayesian manner. Importantly the auditory and visual cues used so far generally consist of single stimuli or objects, rather than complete self-motion-like stimuli. In summary, there is definitely evidence that auditory and visual motion cues are integrated in some-way. For example Calabro et al. (2011) investigated cross-modal enhancement between some features of auditory and visual self-motion cues, and found that the addition of an auditory cue allowed participants to more accurately identify a cue moving independently of the self-motion field. Describing this integration and hypothesising about the underlying mechanism are the aims of this study.

Adding Baysian Priors for deciding whether to integrate

Cheng et al. (2007) wrote that there are three types of integration, depending on when the information was collected. The first is the integration of two current sources of information, the second occurs when

a current piece of information is combined with a prior source. The third is the combining of current and categorical information. Most of the more recent Bayesian integration studies have assumed that the integration is of the first sort: The information available is purely from the current stimuli. This view may need to be extended to take into account priors about the two stimuli, controlled by either our long-term belief about a sense's reliability (e.g. Battaglia et al., 2003; Tassinari et al., 2006) or even influenced by recent emotional stimuli in the modality of interest (Maiworm et al., 2012). The way we integrate must also take into account the associated costs and benefits of integrating the two stimuli. Roach et al. (2006) suggest adding a second layer to the Bayesian integration of cues reflecting the prior likelihood that two stimuli are generated by the same stimulus and need integrating. This addition generalises the rules of integration suggested by Meredith & Stein (1986), that integrating only occurs when the two stimuli are co-located in time and space. This rule of integration may be malleable. In particular, the rules regarding what should and shouldn't be integrated may be adaptable. Shams & Beierholm (2010) describe an 'interaction prior' - a joint distribution which defines how likely any given combination of visual and auditory stimuli are (prior to receiving evidence). They review several works including Ernst (2007) which suggests that one can modify whether two cues will be combined, including those from modalities which are, in both the artificial and natural environments, usually unrelated.

3.1.3 Summary

Besides correctly describing the behavioural response, it is valuable and important to begin linking behaviour with the neural implementation. An example of this is the Normalisation model mentioned previously. This model suggests a method in which plausible neural responses and connections could produce roughly Bayesian responses. Part of this study attempts to apply this model to the behavioural data to generate new predictions which might allow the mechanistic model's results to be distinguished from the descriptive Bayesian explanation. The model was further extended to include some adaptation effects in an attempt to explain intriguing differences in adaptation between modalities.

The initial purpose of the experiment was to develop auditory and visual stimuli for the fMRI experiment to follow. These stimuli needed to be capable of accurate thresholding and needed to be clearly understood by participants. Besides developing a reliable stimulus, the experiment had two key psychophysical areas to investigate: Cue Integration and Perceptual Adaptation.

Integration of a wide range of stimuli appears to be a ubiquitous feature of sensory processing. The precise mechanisms and algorithms underlying this integration are still generally descriptive and incomplete. To investigate how optic flow interacts with auditory cues we developed a new visual stimulus and introduced an auditory motion stimulus in a psychophysics experiment.

Which models of integration would fit the results most accurately? Five models were considered; no integration, Bayesian integration, the Normalisation Model, Bayesian integration with a prior preference for the visual stimulus and a similarly modified Normalisation model. This set includes both models which are descriptive and mechanistic. Although this is not ideal for a comparison it allowed us to validate the results, through the comparison to the Bayesian model while at the same time looking

for ways in which the mechanical model predictions may differ. It was found that the similarity in results between the Bayesian and Normalisation models, and the real data, meant further comparisons with a variety of mechanistic models would not result in useful differences. In the discussion it was suggested that future studies focus on the lowest-coherence stimuli at which the greatest differences between models and data were found.

What are we trying to test? It was hoped that the results would offer significantly more evidence for one model over another. For example, if the Normalisation model fitted the data considerably better than the Bayesian model, then one could conclude that some aspects of the Normalisation model was important in describing how cues are integrated. Similarly if the prior preference for the visual stimulus which was incorporated was found not to be able to explain both congruent and conflict conditions then we would suspect that a different descriptive model would be required (for example a model selecting when integration should occur Ernst (2007)). The models were not chosen specifically with the future fMRI study in mind - although one effect of the Normalisation model, had it been shown to be significantly better at predicting the results than the descriptive Bayesian model, would have been its predictions regarding the weighting of input regions to the integration neurons. This would have provided a strong theoretical reason for expecting to find connectivity changes between regions in the fMRI study.

Which of the questions we could answer about integration depended, in large part, on which sessions were successful. If the conflict condition was successful we would be able to assess whether the visual stimulus captures the auditory or vis-versa. There was concern that the conflict condition would be detected by participants. This in itself would be a useful result in developing the fMRI experiment. The congruent condition also had great uncertainty: Would integration occur and would it be Bayes-optimal?

The adaptation effect can be most easily experienced by fixating near a moving image (such as a waterfall). After a period of time, visual processing in the brain will adapt to this new stimulus and when the observer looks at a stationary object (such as the trees near the waterfall) they will perceive motion in the opposite direction (upwards). One likely hypothesis is that perception is based on the comparison of the responses of neurons preferring upward motion with those that prefer downward motion. Those neurons responding to the downward motion of the waterfall undergo neural adaptation, firing less over time for the same stimulus. When the stimulus is removed, they fire less than their previous average firing rate, and so have a lower firing rate than the remaining (upward) neurons. This results in a perceived upward motion until the neurons re-adapt. If adaptation is taking place in our experiment then after a trial in one direction, the participant's perceived direction will have slightly adapted towards motion in the opposite direction. So we should expect a greater probability of their response to the next trial to be in the opposing direction to the previous trial. Although the trials are brief (less than 4.5 seconds) studies have found adaptation to occur after very brief stimuli (less than 100ms, e.g. Glasser et al. 2011). In summary, we would expect the participant to be less accurate in a trial following an identical trial. For example, if the stimuli are Left, then Left, we would expect the 2nd left to be judged less accurately than in the opposite possibility: Right then Left.

Other hypotheses tested include the compounded increase in the adaptation effect as consecutive trials occur in the same direction. This effect may not be the same in all conditions. For example the

visual stimulus was more constant across the trial while the auditory stimulus varied during the trial, potentially leaving more time for the neurons to re-adapt. A plausible explanation for any adaptation effect was a conscious expectation of trial reversal (especially after several trials in the same direction). One would expect under this hypothesis that the perceived stimuli (presumably those directions recorded in participant responses) would predict the adaptation effect better than the real stimuli. This and other distinctions were tested. Finally, the adaptation effect was modelled by modifying the Normalisation model. The results of the bimodal condition were then compared to the experimentally reported adaptation effect.

The results of this mechanistic model led to predictions regarding the location of the neural adaptation we might expect to find in the fMRI study. Specifically, the model suggested that adaptation would be likely to be restricted to the more sensory regions, rather than the integration region. This led to a testable hypothesis to investigate in the fMRI study.

3.2 Method

3.2.1 Participants

25 participants took part, aged 19-42 (mean 23.4, standard deviation 5.0, 12 female). All reported normal or corrected to normal vision, and none reported hearing loss. Participants were recruited by advertising at the University of Edinburgh.

3.2.2 Apparatus

A sound proof room was fitted with a flat screen monitor (430x270mm), keyboard and headphones (Sennheiser EH2270). Participants wore red-green anaglyph glasses to allow stereoscopic 3D imagery to be used. Participants sat with their faces approximately 600mm from the screen. The virtual horizontal field of view (FOV) was 88°. The actual width of the screen was only 39.4° across. This allowed a clearer sense of forward motion, but, as is standard in 3d simulations presented on computer screens the virtual FOV did not match the actual FOV.

3.2.3 Stimuli

Summary

The stimuli used gave the impression of moving over a plane covered in grey disks. At the same time the subject heard the sound of several nearby objects, all placed at random locations on the horizontal plane, at the same height as the subject's head. The trajectory the participant followed was predefined and followed one of two choices. The direction was completely random each trial. Both choices were forwards, but one was slightly leftwards and the other was slightly rightwards. Because of the large difference between the visual and auditory sensitivity, the auditory trajectory deviated further from directly-ahead than the visual trajectory: The visual stimulus deviated by 4.5°, while the auditory stimulus deviated by 22°. The subjects were instructed to press either the left or right arrow buttons to indicate the direction they thought they were moving in. They were permitted to do this either during or after the stimulus presentation. Due to the difficulty of the task (the coherence was adjusted depending on the accuracy of the participant), subjects generally attended to the whole of each trial before making a decision.

Visual

Visual stimuli consisted of approximately 40 dark grey disks, rendered onto a flat plane, projected onto the screen, superimposed on a grey background (figure 3.1). The disks moved over the surface of the plane, giving the impression of forward motion. Distant dots were masked by a mist or fog, which allowed limited numbers of dots to be drawn. Each dot had a limited lifetime of between zero and one second to stop landmarks being tracked. The dots 'faded' in and out of existence. This meant their addition and removal didn't disrupt the self-motion flow of the stimulus. The centre of expansion was 4.5° to either side of the screen's centre. A black cross was placed in the centre of the screen, slightly

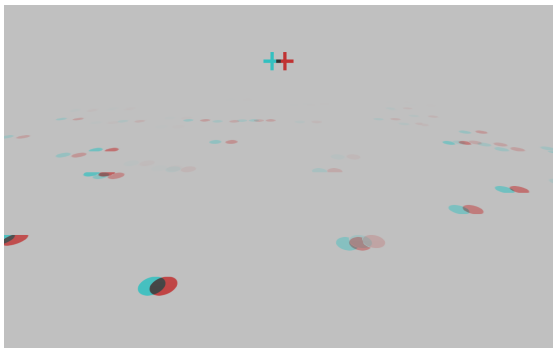


Figure 3.1: An example of the visual stimulus used during the self motion study. See figure 4.7 for the similar stimulus used in the fMRI study in the following chapter. Note this stimulus includes separate red-blue images to provide the 3d anaglyph appearance.

above the horizon, to allow people to keep track of the centre of the screen. Participants were not asked to fixate on the cross, and were permitted to look at any portion of the screen. Note that due to the 3d anaglyph rendering, the disks and cross were shaded red or blue, but appeared black or grey when the two images were combined by the participants.

For this experiment the coherence of the stimuli needed to be varied. A standard method in 2d dot-motion stimuli is to make the majority of the dots move in random directions. When this was tried the sense of forward motion, which we wanted to retain, was lost, unless the task was too easy or the two possible centres of expansion were placed very close to the centre of the screen. Instead, we perturbed the motion of every dot. The direction of each dot remains constant within its lifetime, but when created was sampled from a normal distribution, the width of which was varied to alter the coherence of the stimulus.

Auditory

The auditory stimuli consisted of 4 sound sources; a bus interior, a sewing machine, a small stream, and a waterfall. These were chosen to be quite localised (e.g. a sewing machine rather than rain drops), but not associated with moving objects (e.g. not a vehicle moving past). The intention was that the participant would imagine they were moving past the objects, rather than vice versa. Localised sounds were used because the software performing the auditory processing generated the sounds as point sources. Finally, sounds were chosen which would not attract too much individual attention. We wanted participants to use the whole set of sounds to provide motion information, not just individual sounds.

To generate the spatial component of the auditory stimulus we needed to convolve the sound with a Head Related Transfer Function (HRTF) for every point on its trajectory. An HRTF is an impulse response time course that, when convolved with an auditory cue will produce a new auditory track altered so that it seems to be from the location the HRTF was recorded at. The HRTFs vary between people depending on their head geometry. Alink et al. (2011) and Baumann & Greenlee (2007) used

separate discrete directions to simulate auditory motion, by playing each direction in quick succession, but during pilot trials we found the jumps in direction obvious, distracting and potentially did not provide the motion perception required. However, the HRTFs provided by CIPIC (CIPIC HRTF database, V. R. Algazi, R. O. Duda and D. M. Thompson and C. Avendano), as with all HRTF data sets, only contained recordings for discrete directions. Matlab scripts were written to interpolate between HRTFs to generate an HRTF for an arbitrary direction (available at www.sal.mvm.ed.ac.uk, figure 3.2A). The most obvious difference between a pair of HRTFs for different azimuths is the time difference between left and right ears (the Interaural Time Difference, ITD). If we were to linearly interpolate the values of the two HRTFs, one would find the peaks and troughs would often cancel (see top panel in figure 3.2C). To avoid this, one HRTF was offset to align the peaks and troughs of the two HRTFs. This was achieved by simply cross correlating the two HRTFs and finding the time value with the greatest cross-correlation value. Once aligned the two HRTFs were interpolated as before. The final step was to shift the resulting HRTF an interval the value of which was interpolated between the two original HRTF offsets.

The trajectory of the sound source was divided into points, 50ms apart. To generate apparent motion, segments of the original sound were convolved with HRTFs, each of which was computed for a different point on the trajectory. The sound segment convolved included the neighbouring 11.34ms on either side of the main 50ms segment. These allowed the samples to be mixed together, to avoid any ‘clicks’ or ‘pops’ at the boundaries between segments (Figure 3.2B). Distance was simulated by scaling the sound values by the inverse of the distance. The result of the processing was a new sound track containing the original sound apparently moving along a trajectory, past the listener. The four sound sources were generated in the same way, along parallel trajectories, and mixed together to produce the final track.

To avoid overlearning due to the repetition of the same sound track, 30 different tracks were created, each with the 4 sounds in different locations (always on the horizontal plane, at the level of the subject’s head). The locations of the sounds were chosen to be symmetrical about the direction of motion. The sounds were also positioned so they would be symmetrical about the facing direction at the mid point of the translation. However, during pilot studies it was noticed that some combinations were easier than others. 5 of the most difficult sound tracks were removed, leaving 25 in total. The left and right inputs were switched to simulate motion in either direction (assumes a symmetric HRTF).

To decrease the coherence of the sound stimuli, noise was mixed with the sound track. Originally white noise was used. Some sounds were more easily heard through the white noise than others. The noise was made non-white and used a frequency distribution matching the masked sound. This was achieved by transforming the sound file using the discrete Fourier transform into the frequency domain. The phase of individual components was randomised, and half the components swapped between left and right. The resulting time domain sound was random but had a frequency spectrum which matched the sound track being masked. This made the sound more comfortable to listen to (as it didn’t need to be as loud to mask the cues in the same way), and improved the standardising of the task difficulty across sound tracks.

Initially efforts were made to select the HRTF from the CIPIC HRTF database which best matched the participant. A series of pilot trials were conducted in which participants were required to point

in the direction of the auditory cue using a joystick. No significant difference was detected between the 48 HRTFs tested, suggesting that such preparatory selection would not have any benefit and so the default 'generic' HRTF function was used. Figure 3.3 illustrates an example distribution of the HRTFs illustrating their distribution and the lack of difference in accuracy.

3.2.4 Procedure

Instructions and training

Before collecting experimental data, a series of instructions and training examples were provided to ensure the participants understood the task and to reduce the effect of learning that occurs during training. The introduction began with a description of the environment as a 'foggy plain' containing various sounds the participant would move past. It explained the task, allowing the participant to become used to the stimulus. 8 trials were then given of moderate coherence (88%), which were repeated until the participant reached 7/8 correct. To encourage the use of the sound sources as well as the visual stimuli a series of 16 trials were then presented, in which the visual stimuli had no coherence, leaving the participant to rely on the high coherence (100%) sound sources to determine the direction of motion. Some people responded by reporting the apparent direction of the objects' motion, rather than the simulated self-motion. This was indicated by below-chance accuracy (automatically detected) and the task was explained again, with the emphasis that it is the direction the participant appears to be moving, not the direction the sounds appear to be moving. The session was repeated until at least 15 of the 16 trials were correctly identified. This session was also used as a simple hearing test. If a participant had failed to reach the 15/16 correct criteria they would have been excluded from the study. All participants succeeded in this session after, at most, five attempts.

Staircase procedure

The transformed staircase procedure (Levitt, 1971) was used in this experiment to provide an adaptive coherence estimation procedure. The adaptive procedure allows the limited testing time to be concentrated on those coherences near the threshold of interest. Staircase procedures are simple to implement, and don't make strong assumptions regarding the stability of the psychometric response, unlike the QUEST algorithm (Leek, 2001), which was also considered.

All the staircases used were either one-up-two-down or one-up-four-down. In the former, the coherence would increase if the participant gave one incorrect response, and would decrease if the participant gave two correct responses. One-up-four-down is similar but requires four correct responses for the coherence to decrease (Figure 3.4). In each session these two staircases were interleaved to make it more difficult for participants to predict changes in coherence (Professor Thomas Rammsayer's suggestion, reported in Leek, 2001).

Baseline sessions

Following the instructions and initial training, a series of 'baseline' staircase sessions were presented (step size: 6% coherence). Their purpose was two-fold. First they were introduced to allow participants to practice the task, and so reduce variation in the threshold estimates later. Second, they were used to provide an estimate of the actual thresholds, to allow later session staircases to start near the correct value. The first training session was visual-only (auditory coherence was set at 0%). This session was repeated until a stability criterion was reached. The stability threshold was defined to be that the difference in the average of the coherence in the first half of the session minus the average of the second half had to be less than 6% coherence. There followed an auditory set of baseline sessions, which were for the same purpose as the visual baseline sessions, but with a stability threshold of 11% between the two halves.

Stimuli

The stimuli were presented for a maximum of 4.5s, at which point the screen faded to grey and the software waited indefinitely for a response. The participants could also respond before the end of the trial, which caused the screen to fade to grey and the audio to stop, and the next trial to begin.

Main unimodal experiments

After the baseline sessions, the main experimental sessions were delivered. There were seven 100-trial sessions; four unimodal and two or three bimodal (see figure 3.5). The unimodal sessions were created by reducing the coherence of one modality to zero, but not removing it entirely. Participants were told whether they were in a visual, auditory or bimodal condition before the start of each session. The unimodal staircases always used the threshold estimates from the previous session of the same modality. The main unimodal experimental staircases used a step size of 1% coherence.

Main bimodal experiments

Congruent Condition For the bimodal experiments, the step size was set at one-tenth the difference between the one-up-two-down and one-up-four-down thresholds. The coherence of the two modalities was adjusted in lock-step, so that their unimodal accuracy levels remained equal. During pilot experiments we found that the threshold for the congruent bimodal sessions was much lower than the thresholds in unimodal sessions. Using the coherence threshold discovered in the unimodal sessions as the initial coherence meant most of the session was spent reducing the coherence to the new threshold. To avoid this problem, without needing to repeat the baseline sessions for the bimodal stimuli, an estimate was made as to what the threshold would be. As there was a risk of the data being bias by the initial estimate of the coherence, the stability of the bimodal estimates was monitored in the same way as the unimodal estimates. The stability of the bimodal staircases (mean instability = 1.81) was found to be slightly better than for the unimodal stimuli (mean instability = 1.90), which suggests that this method

had no more bias than the unimodal sessions. The initial coherence values were calculated using the following equations, which were simplified approximations from an earlier, deprecated model.

In the following equations, t_4 and t_2 are the one-up-four-down and one-up-two-down estimates respectively, from the unimodal session. t'_4 and t'_2 are the new estimates for the bimodal session. α and β are temporary variables.

$$\alpha = 2.207 \times (t_4 - t_2) \tag{3.6}$$

$$\beta = t_4 - \alpha \tag{3.7}$$

$$t'_4 = \beta + \left(\frac{\alpha}{\sqrt{2}}\right) \tag{3.8}$$

$$t'_2 = \beta + \left(\frac{\alpha}{1.8285 \times \sqrt{2}}\right) \tag{3.9}$$

Although the basis of the equations was no longer theoretically valid, they provided a useful approximation for the effect of combining both the stimuli to allow the experiment to immediately find the bimodal thresholds, in real time, during the experiment.

The coherence of the two stimuli were changed in 'lock-step' during the congruent condition, so that the difficulty of the two stimuli remained roughly equal. This was partly chosen to provide a simple experimental design, and partly due to the observation by Gu et al. (2008), who observed that the improvement in sensitivity is largest when cues have equal reliability.

Conflict condition For the conflict condition, the two unimodal stimuli were presented, with one having a centre of expansion to the left of the midline, and the other stimulus with a centre of expansion on the right of the midline. One stimulus was set so that its coherence was at the level at which a subject achieved 71% accuracy, while the other stimulus was set to be at the 84% accuracy level. With these coherence values fixed, the direction of the participants' responses were recorded. Fifty trials of each condition were interleaved and presented during the session. The results in this session recorded the proportion of trials the participants selected in the same direction as the visual modality, for each of the two conditions.

3.2.5 Data Preprocessing

Remove first third of trials

Of the 100 trials in an experimental session, 50 were for one-up-two-down and 50 were for one-up-four-down. These interleaved staircases were analysed separately. They consisted of a series of coherence values the participant was tested with. At the start of the session, the stimulus coherence may not have been at the required coherence threshold, so the first few values of the coherence might be bias (either

above or below the actual threshold). To ameliorate this, the first 16 coherence values were removed, leaving 34 values.

Calculate the mean coherence of all runs

The series of coherence values were grouped into sequences of monotonic runs. For the last 34 values of the staircase in figure 3.4 (a one-up-four-down staircase), the runs would be between values 24%-23%-24%-21%-24%-23%-24%. The last value was removed in staircases with odd numbers of runs, such as this, as the estimate can be bias with an odd number of runs (Levitt, 1971). The mean centre point of all runs was calculated, and used as the estimate of this coherence threshold. In this case the estimate was 23.1% coherence.

Remove unstable runs

A staircase might not appear to be settling on a particular value, for instance the first coherence value might be too far from the actual coherence, or the subject may lose concentration during the session. To test whether a staircase was stable a stability criterion was developed during pilot experiments. Six evenly spaced coherence values are chosen from the last 34 values. Their standard deviation, divided by the staircase step size, was calculated. If this value was greater than 3.5, the staircase was considered 'unstable' and discarded. As each subject had two unimodal and two congruent bimodal sessions, a subject was only removed if both of a pair of sessions were deemed to be unstable.

Comparing models

We predicted the expected coherence values using different models for the congruent and conflict bimodal results, based on the unimodal coherence thresholds. The statistical tool used to compare the true bimodal results and the model predictions varied depending on the sessions and models being compared and questions being asked.

3.2.6 Adaptation Analysis

Basic Analysis

To begin the adaptation analysis, each trial was classified into one of four groups in a contingency table, split between those trials which were correct or incorrect and between those following a trial which was in the same or opposite direction. This analysis was computed for each of the three conditions (visual, auditory and bimodal). Simple χ^2 tests were used to test for the adaptation effect. Note that there were slightly more trials in the two unimodal conditions than trials in the bimodal condition. So direct comparisons should not be made between χ^2 values.

Consecutive Trials

The adaptation effect would be expected to increase as more trials in the same direction occur consecutively. Trials were sorted depending on the number of preceding trials known to be of the same direction. The number of incorrect responses in these trials was counted. The proportion incorrect was plotted against the number of consecutive trials. To look for the predicted pattern, the gradient of a linear fit to this graph was calculated, for each stimulus condition. These gradients equate to the proportion of extra trials we would expect the participant to get wrong for each extra trial they had seen of the same direction. Several issues arose with the analysis. We were interested in the effect of additional trials of the same direction, excluding the first trial, but the first large step between the 0th and 1st consecutive trials (already known to experience adaptation) would strongly influence the gradient. Combined with this, the small error bars in this range (due to the large sample sizes) would also bias the result towards a high gradient. To avoid these problems, and to concentrate on the effect of additional trials in the same direction, the linear regression did not include the first data point.

Actual and Perceived Stimuli

One might explain any adaptation by suggesting that the participants are consciously anticipating that the next trial will be in the opposite direction to the previous trial. To put it simply, they may be thinking 'I've had several lefts, must be time for a right'. It should be noted that the completely random selection of directions means this sort of inference was not possible. We tested if the adaptation effect was the same across the three conditions. This was calculated by using a log-linear model for a $2 \times 2 \times 3$ contingency table, testing for block independence between the three stimulus types.[1]

Another way to test whether the adaptation was at the level of sensory processing or decision making was to consider the difference between perception and actual stimuli. If the adaptation occurs at a perceptual or decision making level, we would expect the effect to be stronger if we looked at the participant's direction responses (which we would expect to reflect their conscious perception) and not the actual stimulus directions; If they are consciously compensating for the direction of motion (or if this adaptation is at the level of decision making), one would expect the effect to be considerably stronger when using their perceptions of the stimuli, and not the stimuli themselves.

We could not simply compare the contingency tables of actual and perceived adaptation, as the perceived results suffer from the circularity caused by both axes of the contingency table being dependent on the same data. To avoid this problem, a simple model was created. Each trial in the model was correct with a likelihood set by the experimental results. The effect of adaptation was then included by increasing or decreasing this likelihood by a constant, each trial, depending on whether the previous trial's stimulus was in the same direction. This constant was also set so that the model precisely matched the actual experiment's results. Note that this model only modelled adaptation due to the previous trial.

[1] The validity of the N-way block-independence test (coded in a MATLAB function) was tested using a separate script generating a null distribution of samples. 20,000 null $4 \times 4 \times 2 \times 3$ contingency tables were generated, each containing 1 million samples, in groups of 100. The mean proportion of significant trials (at the 5% confidence interval) in each group was 4.83% ($\pm$0.097% [s.e.]), which lies within the 95% confidence interval of 5%. Based on the lectures notes by Rodríguez (2012).

Finally, the number of correct responses following either trials of the same or opposite directions (or perceived directions) were extracted from the model (estimated by averaging a large number of model iterations). The results were then tabulated, and a χ^2 test or log-linear model used to test for significance.

Adaptation and the Normalisation Model

Although not originally an aim of this experiment, it was realised that the Normalisation Model required only slight modifications to incorporate the adaptation effect. The Normalisation model was modified to include adaptation by dividing the neural responses of the input layers by $1 + \rho r$ where r is the response calculated for the previous trial and ρ is a constant representing the strength of the adaptation. This was chosen rather than subtractive adaptation to avoid negative firing rates occurring. For simplicity the model only used adaptation one-back (i.e. it did not take into account the extra adaptation caused by earlier trials). The value of the constant ρ was fitted from the unimodal adaptation results using an iterative descent algorithm. The two unimodal condition contingency tables were used, and the value of ρ adjusted to fit the model's generated contingency table to the two tables separately. The prediction of interest is the contingency table the model produces for the bimodal condition.

Comparing Adaptation in different conditions

The three conditions had different numbers of trials (due to some sessions being rejected due to instability, etc). To compare adaptation between conditions with different numbers of trials (using χ^2) a simple measure of adaptation was used, independent of the number of trials. To compare the adaptation in different conditions an 'adaptation effect' value was computed for each of the contingency tables, calculated by subtracting the two diagonal cells from the sum of the off-diagonal cells, and dividing by the sum of all the cells. The larger this value the more adaptation will have occurred. This value was reported as a percentage between -100% and 100%. -100% indicates full anti-adaptation, 0% indicates no adaptation and 100% indicates full adaptation.

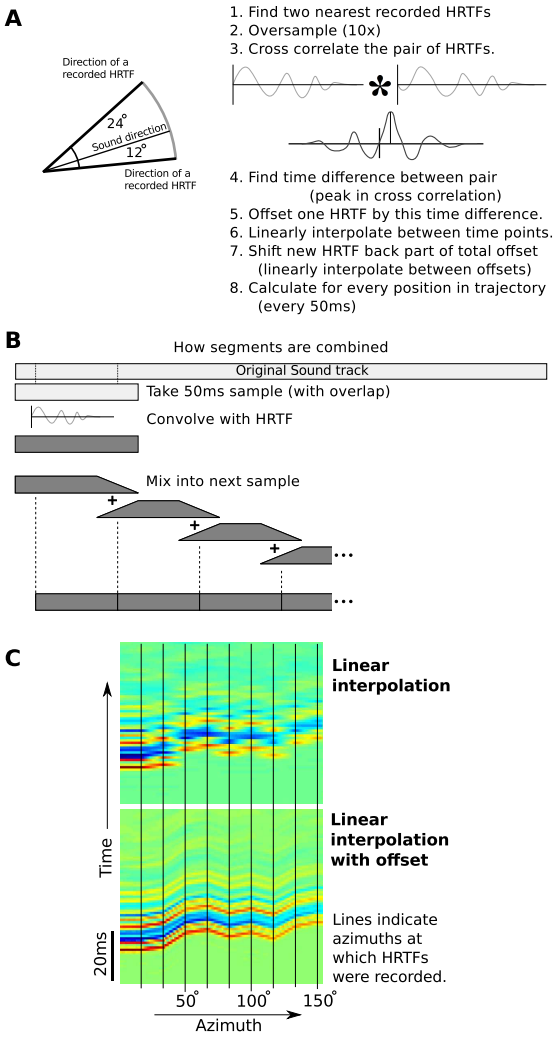


Figure 3.2: AB, Process used to generate auditory stimuli. C, Results with (lower plot) and without (upper plot) offset correction.

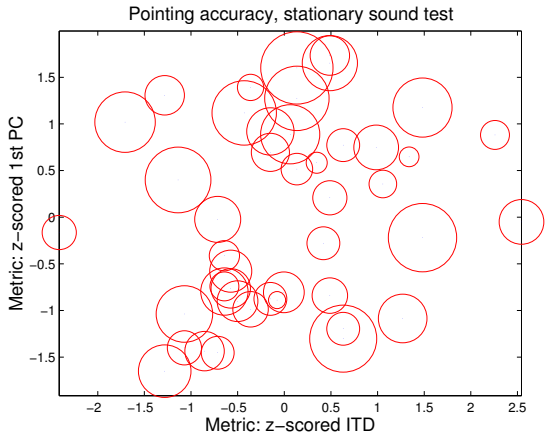


Figure 3.3: Each HRTF was defined by the principle component of the source person's anthropometric information and the HRTF's ITD. The circles indicate the pointing error (larger is a greater error) of a pilot participant.

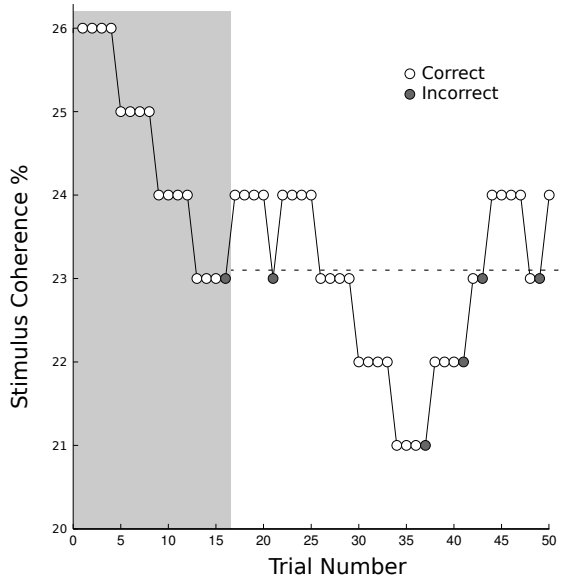


Figure 3.4: Example of the one-up-four-down staircase. The coherence is only reduced when the participant succeeds four times in a row. It is increased if they fail once. In this experiment only the last $^2/_3$ of the staircase is used to estimate the threshold, to ensure the starting values have minimum effect. The dashed line indicates the resulting estimate of the threshold.

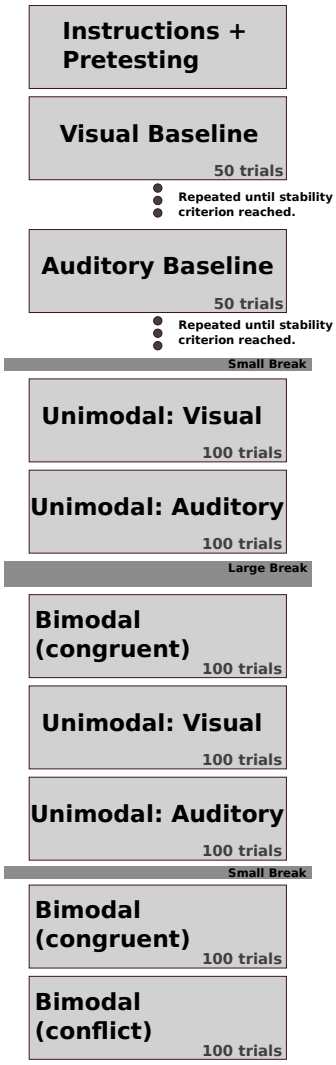


Figure 3.5: Design of the experiment.

3.3 Models and Integration

In addition to the methods described above, several models and methods were used to analyse the results of the experiment. The following section first explains how the Psychometric curve allows more useful parameters of the behavioural responses to be estimated. These are extended through a simple model of a plausible internal representation into how the different cues could be integrated using bayesian integration. Two other models, a modified normalisation model and a visual capture model are described. Finally, how adaptation is related to some of these models is briefly examined.

3.3.1 The psychometric curve

The behaviour of the participants was modelled using a psychometric function. This is a function which describes the proportion of correct responses given a certain stimulus coherence. We decided to use the cumulative normal distribution which is very widely used (e.g. Battaglia et al. 2003) and has some useful mathematical features.

The cumulative normal distribution had two parameters (mean μ_t and standard deviation σ_t) that needed to be fitted to the data. Our experiment provided two points on the curve (t_2 and t_4), the coherence at the points where 70.7% and 84.1% of the responses were correct.

These two accuracy thresholds corresponded to $-0.2168\sigma_t$ and $0.4727\sigma_t$ respectively (taking into account that the floor of the distribution is at 50% and not 0%).

To find σ_t we note that,

$$t_4 - t_2 = 0.4727\sigma_t - {}^{-}0.2168\sigma_t \tag{3.10}$$

which when rearranged,

$$\sigma_t = \frac{t_4 - t_2}{0.4727 + 0.2168} = \frac{t_4 - t_2}{0.6895} \tag{3.11}$$

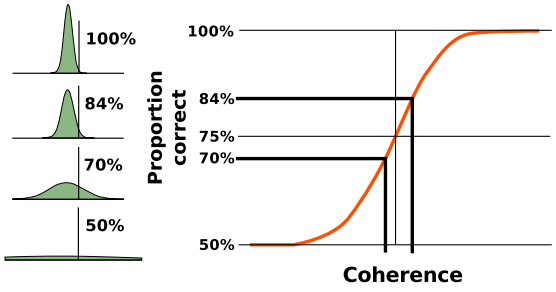


Figure 3.6: The coherence thresholds we were finding (71% and 84%) are the locations where the standard deviation of that source's probability distribution are at certain values. From this we know the variance in an estimate, for a given coherence value.

Similarly, we could find μ_t, as the value of σ_t is known and that μ_t is known to be $0.2168\sigma_t$ above t_2,

$$\mu_t = 0.2168\sigma_t + t_2 \tag{3.12}$$

We used these values to estimate the proportion of correct trials. This accuracy a, for any coherence c, was,

$$a = \frac{1}{2} + \frac{1}{2} cdf(c, \mu_t, \sigma_t) \tag{3.13}$$

Where $cdf(x, \mu, \sigma)$ is the cumulative normal distribution function, summing between $-\infty$ and x with mean μ and standard deviation σ. Note that the $\frac{1}{2}$ offset is because at chance 50% of the trials will be correct (not 0%).

3.3.2 Internal representation

Accuracy and an internal representation

The sensory areas in the brain must, in some way, process information about the direction of the stimuli, and at some point a decision must be made as to whether the centre of expansion is to the left or the right of the midline. To represent this unknown process, it was decided that the perceived direction of motion should be described by a normally distributed random variable.

$$s = N(\mu_s^2, \sigma_s^2) \tag{3.14}$$

The proportion of correct responses corresponds to the proportion of the distribution on the correct side of the midline (the small curves drawn to the left of the graph in Figure 3.6 illustrate how the width of this distribution would change as the coherence of the stimulus varied).

For example if the variance (σ_s^2) of this internal representation was very large, about half of the distribution would lie to the wrong side of zero, indicating a 50% accuracy level. Similarly if the width was very narrow, the response rate would be nearly at 100% accuracy. Figure 3.7 illustrates how the variance of the internal representation is related to the accuracy.

This internal representation's absolute values are unknown, so we fixed the mean response to be minus one, and adjusted the variance to increase or decrease the proportion of the distribution above the zero value 'midline'. So the proportion correct (accuracy),

$$a = cdf(0, -1, \sigma_s) \tag{3.15}$$

Using this value of a, by rearranging equation 3.13 we can find the value of coherence, c,

$$c = norminv(2(a - \frac{1}{2}), \mu_t, \sigma_t) \tag{3.16}$$

We can use the inverse of equation 3.15 to find σ_s, given a. Using this estimate, the coherence can be calculated using equation 3.16.

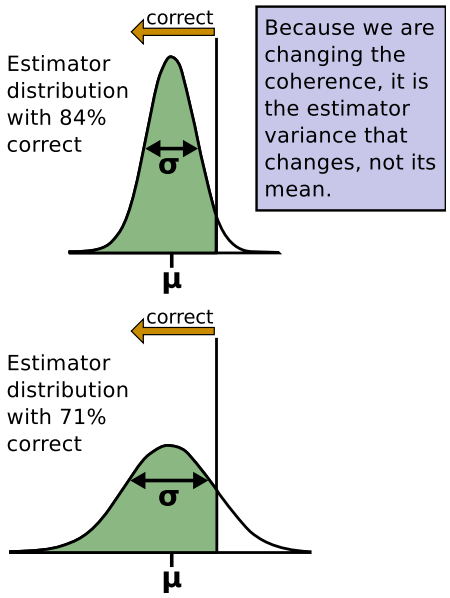


Figure 3.7: The accuracy of the estimates is determined by the proportion of the trials in which the perceived direction of motion lies to the left (correct) of the midline.

Coherence and Accuracy

To summarise; The psychometric curve (described in section 3.3.1) related the coherence (c) to the accuracy (a) of the subject, given their psychometric function's parameters (μ_t and σ_t). The function derived in 3.3.2 related the variance (σ_s^2) of an internal representation to the accuracy (a). These two functions allowed us to calculate the variance of an internal representation (σ_s^2) given the coherence of the stimulus (for an individual psychometric function). Figure 3.8 illustrates this relation. Figure 3.9 further explains the connection.

3.3.3 Bayesian Integration (congruent)

Summary

Consider when both stimuli are presented in a congruent manner (so that they agree) and their coherences are chosen so that in the unimodal conditions they each provide 70.7% accuracy. To find the expect proportion correct in the bimodal condition we must first calculate the standard deviation of the internal representation, in the unimodal condition. This is simply a matter of applying the equations relating the estimator's standard deviation σ_s, derived in the previous section. Solving we find the standard deviation to be 1.8349. This internal representation of the probability distribution of the direction provided by one modality is combined with the other modality (which in this case has the same variance). In the experiment, the situation was inverted; the accuracy was fixed in the congruent trials, and the coherence was reduced. To calculate the expected coherence required, the variance of the unimodal

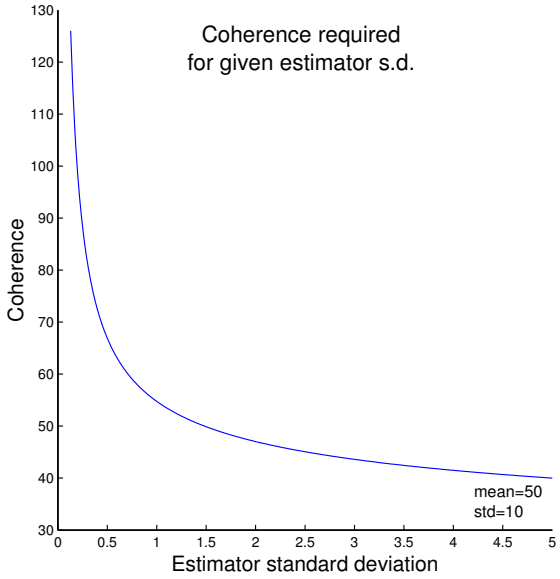


Figure 3.8: Example linking estimator standard deviation with coherence. In this example the psychometric curve had the 75% accuracy point at a coherence of 50 and had a standard deviation of 10. This means at a coherence of 70, the accuracy is 98.75%.

internal representation was doubled. This then cancels out by combining the two stimuli, leaving the same accuracy for the bimodal condition as for the unimodal. Finally, the coherence expected to cause this level of accuracy based on the widened distribution is then calculated.

Example

An example using one of the participant results is described. The participant's unimodal thresholds were $t_4 = 34.425\%$ and $t_2 = 17.475\%$ coherence. Using these two values we estimated the psychophysics function's parameters: $\mu_t = 22.80$ and $\sigma_t = 24.58$.

If the two stimuli are combined in a bimodal session, the coherence required to keep the accuracy the same will be reduced. Specifically, the variance of the internal estimate for the unimodal stimuli would need to be doubled to ensure the variance of the bimodal estimate is the same as the original unimodal variance. The question now is what coherence is required for the unimodal result if the variance of the internal estimate function has doubled. If the estimator function's variance is doubled then the unimodal accuracy will fall from 70.71% to 65.00% (with a similar fall for the 84% accuracy threshold). Given the above mean and variance of the psychophysics function, these new accuracies correspond to two new coherence thresholds: $t'_4 = 24.02$ and $t'_2 = 9.91$. The measured thresholds from the experiment for this participant were: $t_{4bi} = 25.96$ and $t_{2bi} = 12.46$. Note that the measured values have considerable error associated, but appear to match the predictions.

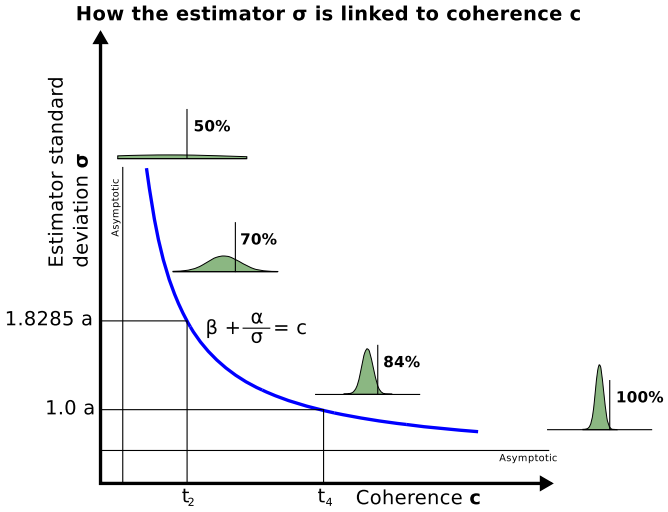


Figure 3.9: To illustrate how the coherence is linked to the estimator standard deviation. For very large coherence values the estimator's standard deviation is almost zero.

3.3.4 Bayesian Integration (Conflict)

We assume that the estimates from the two stimuli are represented as before, as two normally distributed random variables A and V, one with a mean of -1 and one with a mean of $+1$. Their variances are fitted so that the unimodal performances are those measured. I.e. so that:

$$\int_0^{\infty} N(x,-1,\sigma_A)\,dx = 0.7071 \tag{3.17}$$

If the visual stimulus has an accuracy of 70.71% we find that its estimator has a standard deviation (σ_v) of 1.8350. If the auditory stimulus has a unimodal accuracy of 84% it will have a standard deviation of 1.0019.

To combine these estimates, we apply bayesian integration, which finds a new Gaussian distribution with a mean μ_c and variance σ_c^2. These parameters can be found be examining the exponent of the product of the two unimodal Gaussian distributions. This is a standard derivation but is included here for completeness.

$$\mathcal{N}(\mu_c,\sigma_c^2) = \mathcal{N}(\mu_v,\sigma_v^2).\mathcal{N}(\mu_a,\sigma_a^2) \tag{3.18}$$

$$\frac{1}{\sqrt{2\pi}\sigma_c} e^{\frac{-(x-\mu_c)^2}{2\sigma_c^2}} = \frac{1}{2\pi\sigma_a\sigma_v} e^{\frac{-(x-\mu_a)^2}{2\sigma_a^2}+\frac{-(x-\mu_v)^2}{2\sigma_v^2}} \tag{3.19}$$

Ignoring the normalisation term, and just considering the exponent, we can rearrange the second exponent, and equate it to the first:

$$\frac{-(x-\mu_c)^2}{2\sigma_c^2} = -\frac{\sigma_v^2(x-\mu_a)^2+\sigma_a^2(x-\mu_v)^2}{2\sigma_a^2\sigma_v^2} \tag{3.20}$$

Factorising the square terms, and matching the products of 1, x and x^2 across the equality we find the following relations:

$$\sigma_c^2 = \frac{\sigma_a^2 \sigma_v^2}{\sigma_a^2 + \sigma_v^2} \tag{3.21}$$

$$\mu_c = \left(\frac{\mu_a}{\sigma_a^2} + \frac{\mu_v}{\sigma_v^2} \right) \frac{1}{\frac{1}{\sigma_a^2} + \frac{1}{\sigma_v^2}} \tag{3.22}$$

Finding μ_c, we note that we have already defined $\mu_a = 1$ and $\mu_v = -1$, which means the above equation simplifies to,

$$\mu_c^2 = \frac{\sigma_a^2 - \sigma_v^2}{\sigma_a^2 + \sigma_v^2} \tag{3.23}$$

Angelaki et al. (2009) also derives this. Finding the same expression for the new variance as equation 3.21 Quoting Angelaki et al. (2009):

> According to [this equation], the largest predicted improvement in sensitivity (a decrease in the variance of the bimodal estimate by a factor of $\sqrt{2}$) occurs when the two cues have equal reliability (i.e. $\sigma_1 = \sigma_2$). In the extreme case where one cue is much more reliable than the other ($\sigma_1 << \sigma_2$), behaviour is captured by that cue such that $\sigma_{BIMODAL} \approx \sigma_1$.'
>
> *(Angelaki et al., 2009).*

This is the one of the reasons that equal coherences in the congruent bimodal sessions were used. By making the two modalities equal we would hope to generate the maximum difference between the Bayesian- and no- integration models. This is also the explanation for the earlier assertion, that the variance of the internal unimodal representations would need to double for the bimodal condition to have the same accuracy as the original unimodal conditions.

Putting in the values for the estimate of the mean above, gives estimates of the internal distribution's parameters,

$$\mu_c = \frac{1.8350^2 - 1.0019^2}{1.8350^2 + 1.0019^2} = 0.541 \tag{3.24}$$

$$\sigma_c^2 = \frac{1}{\frac{1}{1.8350^2} + \frac{1}{1.0019^2}} = 0.7733 \tag{3.25}$$

Using the earlier equations in section 3.3.2 we find the expected accuracy a_e to be,

$$a_e = \int_{x=0}^{\infty} N(x, \mu_c, \sigma_c) = 0.7308 \tag{3.26}$$

So we would expect 73.1% of trials to be in the direction of the most coherent stimulus.

3.3.5 Normalisation model

Summary of the Model

Ohshiro et al. (2011) suggested that a form of divisive normalization could be used when integrating multiple cues to account for how optimal bayesian integration may be computed at the neural level. They describe a model in which two layers of unimodal primary neurons respond to two sensory modalities. The responses from these unimodal neurons feed into a layer of multisensory neurons. Each neuron in this multisensory layer combines responses from those unisensory inputs with matched receptive fields.

At the multisensory stage, each input has a nonlinearity applied (e.g. caused by normalisation or synaptic depression). The sum of the relevant unimodal inputs is calculated and transformed by another non-linear transform. This one is an expansive power-law function, representing the transformation from membrane potential to firing rate. The input is then divided by the weighted sum of all the other multisensory outputs.

Modification of model for this experiment

Overview The key addition to the model described is the generation of the unimodal inputs. The assumption made is that the stimulus will appear to be from a direction sampled from a normal distribution, centred on the correct direction of motion. The width of this distribution is set so that it reflects the proportion of correct unimodal trials.

Unimodal sensory neurons The first step in the model is the activation of the unimodal sensory neurons. In this model each neuron has been given a Gaussian receptive field, based on the model by Ohshiro 2011 (online methods). The choice of a gaussian receptive field by Ohshiro presumably reflects the shape of most tuning curves reported in the literature and the choice used over several decades when modelling neural systems (e.g. Kaplan et al., 1979). The neurons respond to a range of centres of expansion (Figure 3.10B). The peak of the neurons' receptive fields are spread evenly over the range of possible directions. The width of their tuning curves is equal. This width is an important parameter, which will be discussed later.

What is the stimulus? To model the uncertainty in the stimulus, the direction of the perceived motion, d, is sampled from a Normal distribution, with a mean of one, and a variance dependent on the coherence of the stimulus (figure 3.10A). A lower coherence causes a higher variance, and a greater likelihood of a value less than zero being chosen (i.e. a wrong choice). The variance is chosen so that the accuracy for the unimodal sensory neurons is equal to the threshold of interest (e.g. 70.71%).

Sublinear response to stimulus intensity The above model gives a response from each neuron to a given stimulus, a different response each trial, depending on the perceived direction, d. The next stage of processing is for *'the unisensory inputs [to] increase monotonically, but sublinearly with stimulus intensity'*(Ohshiro et al., 2011). To model this, the activity of the sensory neurons is transformed by the function $log(x+1)$.

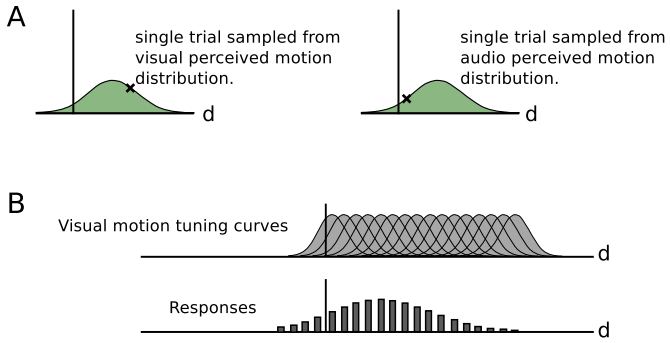


Figure 3.10: A. Sampling from visual percept distribution and from audio percept distribution. B. The tuning curves and responses of the visual CoE neurons.

Before this transform, the inputs are multiplied by the coherence, reflecting the multiplication with the stimulus intensity described in the paper. Note that this has no effect in the congruent case, as both stimuli have the same coherence.

Bimodal Neurons: Integration The bimodal neurons' responses are calculated as described in the paper. First a weighted linear sum of the input neurons is calculated:

$$E_j = w_1 I_1(x_j) + w_2 I_2(x_j) \tag{3.27}$$

Where $I_1(x_j)$ and $I_2(x_j)$ are the inputs from the two unimodal neurons of different modalities. w_1 and w_2 are the weights assigned to each modality. Then this is normalised:

$$R_j = \frac{E_j^n}{\alpha^n + \left(\frac{1}{N}\right) \sum_{k=1}^{N} E_k^n} \tag{3.28}$$

The response R is calculated for every bimodal neuron, each taking its inputs from two input neurons with the same tuning curve centre (of different modalities). The two parameters, n and α, in the above equation need to be fixed. We use the same values as suggested by Ohshiro et al. (2011), $n = 2$. $\alpha = 1$. To read out the perceived direction, the integrated result is defined to be the preferred direction of the bimodal neuron with the highest firing rate. To get a measure of accuracy, this model is repeatedly applied, with inputs sampled from the normal distribution described above. The accuracy is the proportion of trials with an estimate greater than zero.

Model results Figure 3.11 shows the responses of three populations of neurons; two unimodal populations (blue and green) and the bimodal integration neurons (dashed black). The figure illustrates how the response of the integration neurons is both super-additive and combines the two population inputs to provide a better estimate of the stimulus direction.

This behaviour depends critically on the width of the tuning curves of the individual unimodal sensory neurons. If the neurons have much narrower tuning curves, integration does not fully take

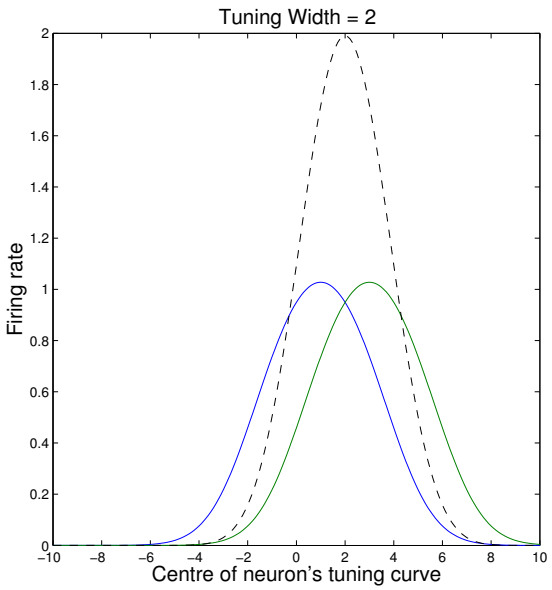


Figure 3.11: Response of unimodal neurons (blue and green) and the response of the bimodal integration neurons (dashed black). Each neuron is a different point on the line, with the x axis indicating the preferred centre of expansion of the neuron. In this example, the tuning curve s.d. = 2.0.

place, and the response becomes suboptimal. Figure 3.12 illustrates such a situation.

This leaves a final question, what is the width of the tuning curves? Qualitatively, the response of the subjects to the conflict condition provides some initial insight: Some of the subjects said they noticed that a few of the trials had conflict. This ability to detect a conflict condition suggests that integration was not taking place, but the only occasional detection also suggests the width is almost always wide enough to enable integration. We assume from this that the tuning curves generally integrate successfully, with only occasional failures. Figure 3.13A illustrates that the normalisation model with wide tuning curves (width=5) fits very closely the results obtained with Bayesian integration. With slightly narrower tuning curves (std=2) the accuracy at the lower-coherence end of the graph begins to drop (Figure 3.13B). The sampling from the two unimodal stimuli distributions are within 2 of each other roughly 99% of the time (given a 71% unimodal accuracy). For the same unimodal accuracy (71%) the two sampled points are within 0.6 of each other only 57% of the time, meaning that integration will not occur very frequently. Figure 3.13C shows the result with a tuning curve of only 0.6. The suboptimal performance at the lower coherence levels reflects the lack of integration (due to many responses such as those in figure 3.12). As a compromise between these extremes, a tuning curve width of three was chosen. This value is relatively arbitrary, and would require an additional experiment to more precisely determine its value.

The tuning curve widths discussed above are given unitless values. To understand why, we must recall that the direction of the stimulus is selected from a Normal distribution with a mean of one, and a variance dependent on the coherence of the stimulus. To allow real angles to be assigned to these values, we can consider first the visual stimulus. The centre of the normal distribution that the visual

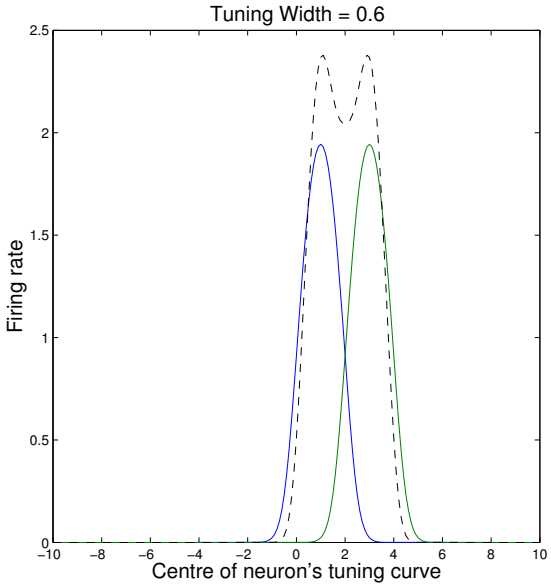


Figure 3.12: Response of unimodal neurons (blue and green) and the response of the bimodal integration neurons (dashed black). Each neuron is a different point on the line, with the x axis indicating the preference of the neuron. Tuning curve = 0.6.

stimulus is sampled from, is 3.5° from the centre. In the unitless 'percept' part of the model this is equivalent to the distance of one unit. So we can determine that the standard deviation (chosen to have a standard deviation of 5) of the receptive fields in the visual modality is $3.5^\circ \times 5 = 17.5^\circ$. For the auditory stimulus (with a much wider angle of 22°) the standard deviation is equivalent to 110°. These values may seem very wide, but if they were much narrower (e.g. just $7^c irc$ and $44^c irc$ respectively) the likelihood of integration would be greatly reduced.

The final question that needs asking is, are these tuning curve widths physiologically plausible? Precise tuning curves of MST neurons for different directions of motion have not yet been established. However, Duffy & Wurtz (1995) report that most expansion/contraction neurons in MST were restricted to respond more in three or fewer of the nine centres-of-expansion that they tested (in a $90^\circ \times 90^\circ$ grid). This suggests a tuning curve widths of between 30 and 60°. The gaussian with a standard deviation of 17.5° has a FWHM of 41°, which is broadly of the same width as the reported range in MST. The study by Gu et al. (2008) found the width of the tuning curves for both visual and vestibular responses to consist of very wide angles. It is important to note that it may be the gradients of the edges which give information about the heading direction and not the peak of the tuning curve. Critically it is not necessarily true that narrow tuning curves are the most precise for a population code (e.g Pouget et al., 1999).

No literature exists on the receptive field of centre-of-expansion for auditory cues. However, some experiments have looked at the receptive field for auditory localisation. Middlebrooks & Pettigrew (1981) found that in cat primary auditory cortex, neurons had receptive fields of between 20° and 180°

across, most of about 40° across. It is very difficult to extrapolate from this to predict the receptive field size in human auditory-centre-of-expansion tuned cells. One difference that needs mentioning is that humans are unable to rotate or move their pinnae. Also it is unclear how to move from the receptive field of localisation to the receptive field of expansion. One might reasonably conclude however that the auditory motion receptive field is larger than that of visual motion. The model standard deviation of 110° gives a FWHM of 257° which is comparible in scale to the large receptive fields reported here. For comparison V1 receptive fields (outside of the fovea) are of the order of 5 degrees, so a factor of about 8 times smaller than the equivalent auditory receptive field. This is close to the ratio of the model's visual:auditory receptive field standard deviations.

Clearly the choice of tuning curve width is slightly arbitrary. Reassuringly, while testing this, it was found that a fairly broad range of tuning curves had similar results, especially as one increased the curve width. Other features of neural activity are absent in this model. First their individual responses are actually somewhat stochastic and consist of discrete action potentials, rather than continuous varying firing rate values. Second, some neurons in MT and MST respond to all directions of motion, not just motion in the neuron's preferred direction (Duffy & Wurtz, 1995).

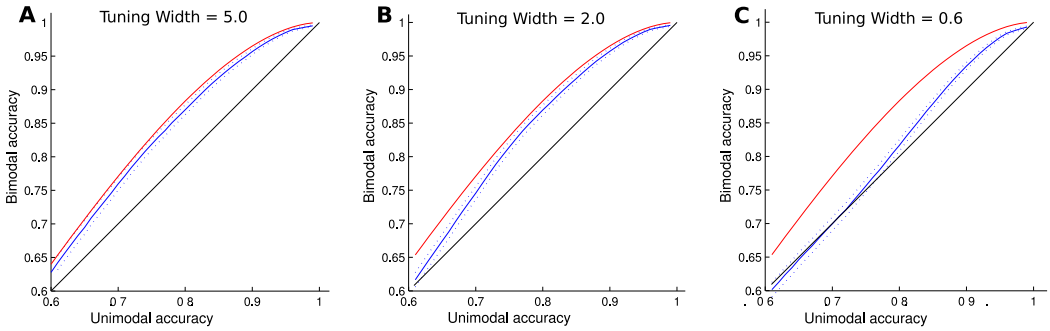


Figure 3.13: The predicted bimodal accuracy given unimodal accuracies, for three models. In red: Bayesian integration, assumes the percept distribution is narrowed by $\frac{1}{\sqrt{2}}$. In black: no integration, the accuracy is the same as for the unimodal case. In blue: the predicted accuracy of the normalisation model (dotted lines indicate 95% CI). The stimuli sampling distribution, used to select the perceived unimodal direction each trial, has a mean of one, and a standard deviation which varies depending on the unimodal accuracy (e.g. 0.55 in the 71% accuracy condition). The tuning curves have a width of A. 5, B. 2 and C. 0.6.

3.3.6 Visual Capture

Visual capture (or the 'ventriloquism effect') is the reported dominance of visual information over other modalities in the creation of a percept (Mateeff et al., 1985). In general, if complete visual capture occurs, one would expect the auditory stimulus to have no effect on the accuracy or response direction of the participants, in the bimodal sessions. For example, in the conflict condition, one would expect the

participant to chose the visual direction with a frequency of either 71% or 84% of trials (depending on the condition). In the congruent sessions one could expect the accuracy to be unchanged from the visual unimodal sessions.

In the earlier description of Visual Capture, the complete capture of the stimulus by the visual modality was described. However, such extreme visual capture may be unlikely. Instead, more subtle effects may occur. Battaglia et al. (2003) found visual capture even in stimuli of the same coherence. We might expect that, in the congruent sessions the accuracy may be slightly less than that predicted by the bayesian integration model.

In the conflict session, without 'capture', the number of responses in the direction of the high coherence stimulus should be unaffected by which of the stimuli are high-coherence. If visual capture is invoked, one would expect the participant to select the direction of the low-coherence stimulus more if it is a trial in which that low coherence is visual.

A related issue is binding of the two modalities. In the conflict condition, integration or the 'capture' effect may not occur due to the large difference in the stimuli - indeed as is reported in the results section, many of the participants were able to detect the conflict condition. This conscious detection does complicate the interpretation of the conflict condition results, and the lack of binding may explain features in the results which one might attribute to visual capture.

Bayesian Interpretation

Visual capture could be expressed as a prior preference for visual cues, over auditory cues. Below is a brief mathematical explanation. This may be unnecessary for the purposes of model comparison due to the lack of detailed evidence for one model or another in the data, and uncertainty in the probability distribution of the prior. However, it may be a useful framework for future experiments investigating these effects. We note that a generative model of the direction of a visual stimulus, $p(\mathbf{x}|\theta)$ describing how data regarding the direction of the visual stimulus $\mathbf{x}$ is generated by the model can be used to estimate the parameters of the model (θ):

$$p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{p(\mathbf{x})} \tag{3.29}$$

Where $p(\theta|\mathbf{x})$ is the estimate of our parameters given data, $p(\mathbf{x})$ is the probability of the data, which can be ignored as it is equal across the comparison of parameter values. $p(\theta)$ is a prior estimate of the model's parameters. For this example it is our prior belief about the accuracy of the two modalities which is of particular interest. Specifically the variance σ of the data $\mathbf{x}$. We assume that the data is normally distributed, around a mean μ and with a variance of σ^2. We assume the datapoints are conditionally independent (given θ) The denominator remains equal for all values of θ and so can be ignored.

$$p(\theta|\mathbf{x}) \propto \prod_{i=1}^{N} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} \; p(\theta) \tag{3.30}$$

The step which varies from previous derivations of this expression is the assumption that the prior

on θ is flat. In our case, we hypotheses that people have a bias and assume that the variance of the visual stimulus will be smaller than the auditory.

Leaving aside the decision of which function to use to represent the prior expectation over σ, we continue with estimating the standard derivation, first switching to log-likelihoods and differentiating with respect to our parameter of interest (in this case σ).

$$L \propto -\sum_{i=1}^{N} log\ \sigma - \sum_{i=1}^{N} \frac{1}{2} log\ 2\pi - \frac{1}{2} \sum_{i=1}^{N} \left(\frac{x_i - \mu}{\sigma} \right)^2 + log\ p(\theta) \tag{3.31}$$

$$\frac{\delta L}{\delta \sigma} \propto -\sum_{i=1}^{N} \frac{1}{\sigma} + \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{\sigma^3} + \frac{\delta}{\delta \sigma} log\ p(\theta) \tag{3.32}$$

Setting equal to zero, and rearranging, we find:

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N} + \frac{\sigma^3}{N} \frac{\delta}{\delta \sigma} log\ p(\theta) \tag{3.33}$$

To simplify, using s^2 to represent the true variance of the data (i.e. without the prior),

$$\sigma^2 = s^2 + \frac{\sigma^3}{N} \frac{\delta}{\delta \sigma} log\ p(\theta) \tag{3.34}$$

We can see our estimate of the variance is equal to the actual variance plus a complicated term introducing the effect of the prior, $p(\theta)$. What prior could we use? There are a large number of choices. For simplicity, I've chosen to use an exponential decay, indicating that we would expect the variance to be nearer zero than far from zero (i.e. visual capture's expectation that visual uncertainty will be low):

$$p(\sigma) = \frac{1}{a} e^{-\sigma / a} \tag{3.35}$$

Substituting in, we find that:

$$\sigma^2 = s^2 - \frac{\sigma^3}{Na} \tag{3.36}$$

The division of the last term by N indicates that with more information, our estimate of σ^2 should depend less on this prior, and more on the data, from s^2 (Figure 3.14). We don't have access to a direct measure of N, so we simplify the equation by removing N, as this is the same across trials, and letting the parameter $A = aN$ include this element of the distribution. This leaves the one parameter, A, which is a constant reflecting how much bias to apply (a smaller number means we have a stronger preference for smaller variances). Using the same equation (3.23) as before, we can substitute a different estimate of the variance of the visual stimulus.

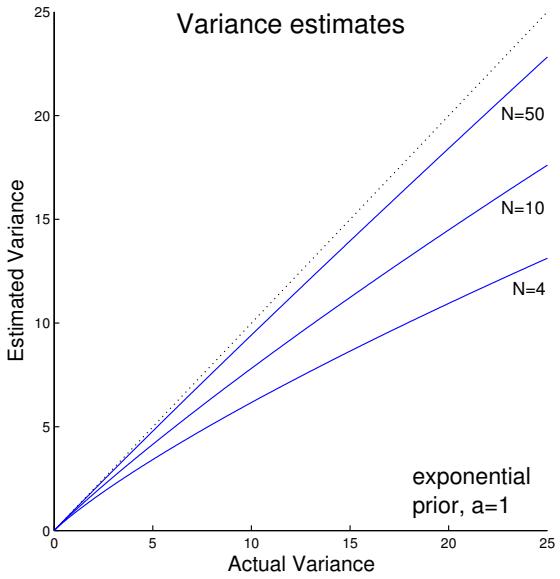


Figure 3.14: How the estimated variance will vary with the actual variance for different values of N. The prior emphasises smaller variance. As the number of samples increases we can depend less on the prior and more on this extra data. Note that we don't use a value of N in further analysis.

3.4 Results

3.4.1 Sound Stimuli

Because the auditory stimuli were restricted to 25 soundtracks, it was of interest to determine if some of these soundtracks were easier than others. The proportion of correct guesses for each soundtrack was computed and compared between participants using an ANOVA to test whether there were significant differences between the soundtracks. Even with the most difficult soundtracks having been removed, there was enough variation between samples to cause a significant variation in their difficultly (f = 4.98, $p < 0.001$, mean score = 22%, s.d. = 6%). This sort of variation may hold in other stimuli (such as translational dot motion) but hasn't been detected yet, as dot motion stimuli are all unique.

3.4.2 Stability

To process reliable, good quality data, those staircases with poor stability were removed from the experiment. The stability of a session was calculated by selecting seven evenly spaced threshold points from the session (including the first and last trial), calculating the standard deviation of these seven values and dividing this by the step size of the session. A threshold value of 3.5 was chosen during pilot studies as a reasonable threshold at which the graph of coherence subjectively appeared to be deviating too far from a normal stable trajectory. Above this threshold the session is rejected.

In those cases where two staircases were removed for the same measurement, the participant was dropped from the analysis. The stability threshold meant that of the 25 participants, 4 were rejected due to low stability. To summarise, for unimodal visual motion, 5 staircases (of 100) were removed due to low stability (rejecting 1 participant). For unimodal auditory motion, 8 staircases were rejected (rejecting 2 participants), and for the bimodal sessions, 4 staircases were rejected (rejecting 1 participant). One more participant was removed due to staircases in which the coherence remained at 100% (suggesting they were unable to perform the task, possibly due to poor attention or understanding).

A further two participants were rejected when it was found that their unimodal thresholds were reversed (so that the 84% threshold was lower than their 71% threshold). This reduction left 18 participants in the study.

3.4.3 Unimodal Thresholds

Figure 3.16 shows the thresholds for one-up-two-down and one-up-four-down auditory stimuli for the 20 participants (including the two rejected due to reversed thresholds). The variation in threshold between participants was quite striking. To confirm that this variation represents inter-subject variability and not just variance in the actual estimate, the two thresholds from the two unimodal sessions were plotted against each other (shown in figure 3.15). The correlation between the sessions (across subjects) indicates that the majority of the variation is due to inter-subject differences, and not noise within the measurement. Partly though, this correlation is due to the starting values of the sessions being based on

Stimulus	Accuracy	Correlation	p-value	C^2
Auditory	71%	0.871	3×10^{-5}	76%
Auditory	84%	0.938	3×10^{-8}	88%
Visual	71%	0.950	2×10^{-9}	90%
Visual	84%	0.973	5×10^{-11}	95%

Table 3.1: Correlations between unimodal sessions, testing for their stability.

the results of the initial training and calibration sessions.

On average, across subjects, 87% of the variation between the values of the second session are explained by the variation between the values of the first session (Table 3.1).

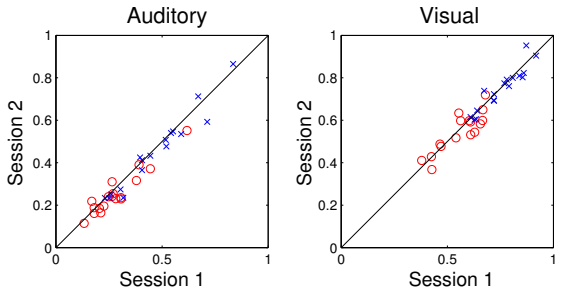


Figure 3.15: Plotting the unimodal thresholds from two sessions against one another, to test for stability. Only subjects with both sessions above the stability threshold were plotted. Circles are the 71% accuracy thresholds, crosses are the 84% accuracy thresholds.

3.4.4 Bimodal Conflict Detection

Of the 18 participants who reached the stability criterion (and weren't rejected due to reversed thresholds), 2 did not complete the conflict condition due to time limitations or fatigue. A serious issue in this experiment was that most of the participants (10 of these 16) detected the conflict between stimuli. These results were determined by asking all participants after their experiments, whether they 'noticed any differences in the last session' (potentially a leading question). The proportion of responses towards the visual stimulus in the conflict condition were tested between these two groups, but no significant difference was found (t-test between groups. visual coherence greater-than auditory coherence, means: undetected, 0.76 (n=6), detected, 0.69 (n=10), p = 0.32 [t=1.03, df=14]. visual coherence less-than auditory coherence, means: undetected, 0.45 (n=6), detected, 0.37 (n=10), p = 0.41 [t=0.85, df=14]).

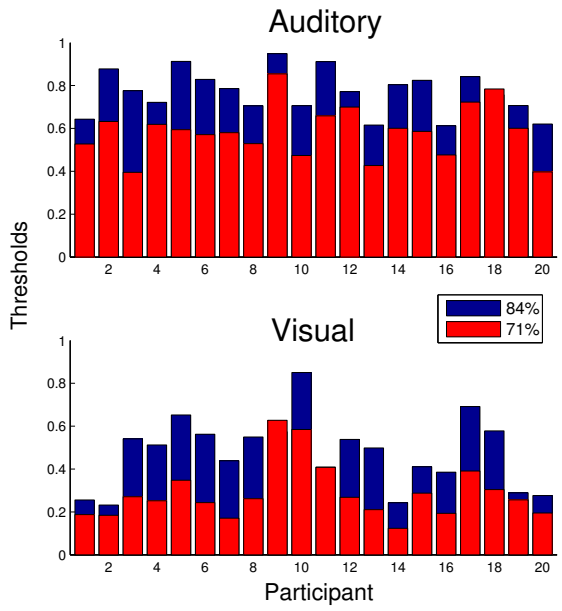


Figure 3.16: The unimodal responses to the two stimuli (included are the two rejected participants, 9 and 18).

3.4.5 Congruent Bimodal Integration

Bayesian Model

The top two graphs in figure 3.17 show the actual results and those predicted by the Bayesian model. Many appear to lie on, or very close to the line, with the Bayesian model's predictions being insignificantly different to the actual values (Table 3.2).

Comparing these using the likelihood ratio test, finds the Bayesian model predicts the results significantly better than the model with no integration (log-likelihood ratio = 30.7. The threshold at the $p < 0.001$ level is 18.5).

Normalisation Model

The lower two graphs in figure 3.17 show predictions of the Normalisation model. It appears there is a trend for the Normalisation model to fit the data somewhat better than the Bayesian model. We found that, comparing between models, the normalisation model was not significantly more accurate (log-likelihood=0.95, threshold=9.87 [2]). The slight improvement appears to be because the Bayesian

[2]To determine whether this trend is significant we first used the 'Bayesian t-test' (Wetzels et al., 2009). This allows us to report the evidence for and against each model, and express the ratio of probabilities as a Bayes Factor:

$$BF_{fromBAYES/\overline{fromBAYES}} = 2.39 \tag{3.37}$$

$$BF_{fromNORM/\overline{fromNORM}} = 5.60 \tag{3.38}$$

However, to compare between the data produced by the two models we could not use the comparison above as the two sets of data (the error values between predictions and actual data) were different (which would mean the denominators would differ).

Model	Condition	t	p
Bayesian	Visual High Coherence	2.01	0.06
	Visual Low Coherence	-0.54	0.59
	Auditory High Coherence	1.35	0.19
	Auditory Low Coherence	-1.51	0.15
No integration	Visual High Coherence	-4.48	3.27^{-4}
	Visual Low Coherence	-7.04	1.99^{-6}
	Auditory High Coherence	-5.69	2.6×10^{-5}
	Auditory Low Coherence	-9.12	5.8×10^{-8}

Table 3.2: t-test results of the comparison of the models with the real results. Bayesian and No-integration models with Frequentist results.

model overestimates the accuracy of integration in the low coherence condition, while the Normalisation model predicts a slight reduction in accuracy in the low coherence condition (Figure 3.17).

In summary, the two models both do almost equally well at estimating the bimodal thresholds, but they both do far better than assuming no-integration is taking place (paired t-test on distance from the unimodal thresholds $p < 10^{-6}$).

3.4.6 Conflict Bimodal Integration

The main competing models available to test the conflict condition session, are the Bayesian integration model and visual capture. Battaglia et al. (2003) report that visual capture occurs, even when stimuli have been matched for accuracy in unimodal sessions. Our experiment finds a similar result.

Figure 3.18 illustrates the conflict session. Half of the one hundred trials were with the visual coherence higher than the auditory and the other half the reverse.

Visual Stimulus more coherent

In those trials in which the visual stimulus was the most coherent, the response to the stimulus was, averaged across subjects, the accuracy predicted by the Bayesian model (p=0.71). There's a sizeable spread of responses, significantly greater than one would predict (sampling from a binomial sampling with N=50, p=0.73 suggests 95% of the samples would lie between 63% and 82% - but half lie outside this range). The distribution also is very non-normal (Kolmogorov-Smirnov to the normal distribution, $p < 1 \times 10^{-4}$ in both the high visual and high auditory coherence conditions), with an interesting cluster in the high visual coherence condition around the 84% accuracy level. This is the accuracy that would be

Instead we used the likelihood ratio test and used the Neyman-Pearson lemma to derive the 5% significance threshold. Which for this ratio is $norminv(1-\alpha, 0, \sqrt{n})$ (where norminv is the inverse cumulative normal distribution). Note that only the visual thresholds were used in this calculation, as the auditory and visual sets of data are highly correlated due to the data being from the same participant, and uses the same model prediction inputs and the same bimodal results.

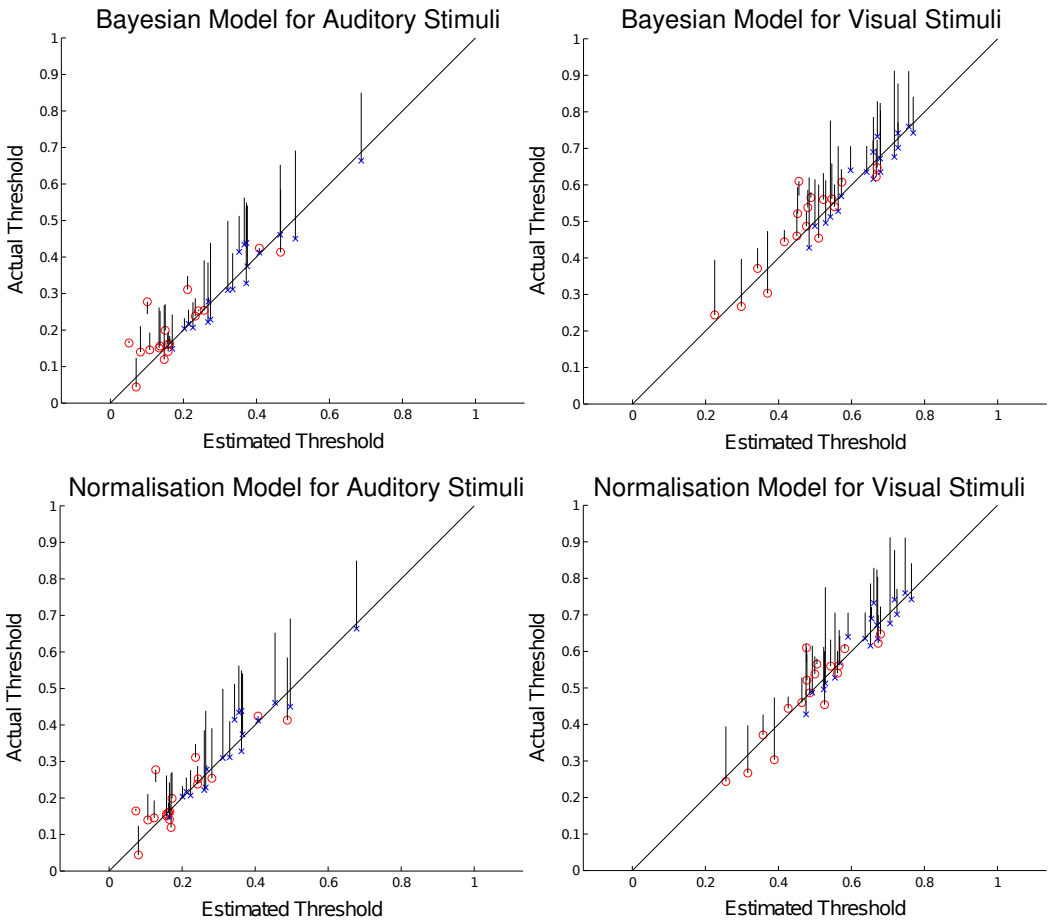


Figure 3.17: The results of both the Bayesian and Normalisation model. The symbols indicate the bimodal results, while the lines indicate the difference between these results and the unimodal results. Blue crosses are at the 84% thresholds, while red circles are at 71% thresholds.

expected there if the participants ignored the auditory cue completely. It also appears that the auditory greater-than visual coherence condition has more variance, but this is not a significant effect (standard deviation of the two groups: 0.13, 0.18, f-value: 1.96, p = 0.20).

Auditory Stimulus more coherent

In the high auditory condition the data was significantly different from the Bayesian estimate. Most subjects had on average, bias their responses towards the visual direction of motion ($p < 10^{-5}$). Even looking at just those subjects who did not report detecting the conflict condition we found this to be significant (p = 0.02, two-tailed t-test, t = 3.20, dof = 5).

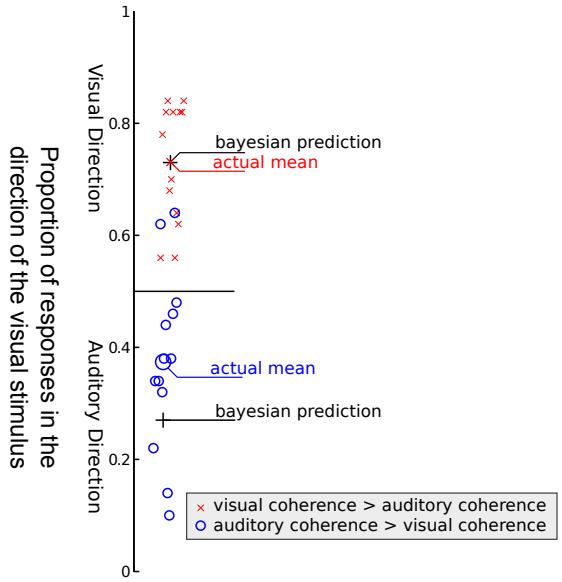


Figure 3.18: The unimodal responses to the two stimuli (included are the two rejected participants, 9 and 18).

3.4.7 Bayesian Visual Capture

Outline

The apparent visual capture, described above, is potentially that predicted by the modification of the bayesian function to include a prior preference for visual cues instead of auditory ones (section 3.3.6). In the conflict condition with a greater auditory coherence than visual coherence, the average proportion of responses given towards the visual direction was 40.37%. We can use this to estimate the key parameter A of the new model, which describes the model's modification to the estimate of the visual stimulus' standard deviation (σ_v). We assumed that this would be reduced below the default standard deviation, as we're suggesting that the error is underestimated in the visual modality.

Without this bias, we assume that at the 71% accuracy point, the internal representation of the stimulus will have a standard deviation of 1.8350. The proportion in the direction of the visual stimulus (40.37%) can be used to estimate the actual value of this parameter (given the standard deviation of the auditory estimate remains unchanged).

Deriving the new scores

Using equation 3.23 and the cumulative normal distribution (equation 3.26) one can calculate the value of the visual estimate's standard deviation which fits the actual number of correct responses to the low-visual coherence condition most accurately. For this data, this value is $\sigma_v = 1.2122$, instead of the original estimate of 1.8350.

Finally, we rearranging equation 3.36 to find the parameter A which describes the prior for visual

stimuli:

$$A = \frac{\sigma^3}{s^2 - \sigma^2} = \frac{1.2122^3}{1.8350^2 - 1.2122^3} = 1.1231 \tag{3.39}$$

Using the calculated value of A above, we can solve equation 3.36 to find σ_v, the standard deviation of the visual estimate when the visual stimulus has high coherence.

What effect does this have on the congruent estimates?

To find the estimate of the congruent thresholds we need to find the new weightings for the two distributions, and the resultant, combined, variance. The standard deviations estimated from the conflict condition (in the low visual coherence condition) were: $\sigma_a = 1.8350$, $\sigma_v = 1.2122$.

To find the weights we use the same equation as in section 3.3.3:

$$w_a = \frac{1/\sigma_a^2}{1/\sigma_a^2 + 1/\sigma_v^2} \tag{3.40}$$

But this time substitute in our new values of σ_v:

$$w_a = \frac{1/1.8350^2}{1/1.8350^2 + 1/1.2122^2} = 0.3038 \tag{3.41}$$

and

$$w_v = \frac{1/1.2122^2}{1/1.8350^2 + 1/1.2122^2} = 0.6962 \tag{3.42}$$

We can solve the same equations for the standard deviations in the high visual coherence condition, giving: $w_a = 0.3722$ and $w_v = 0.6278$.

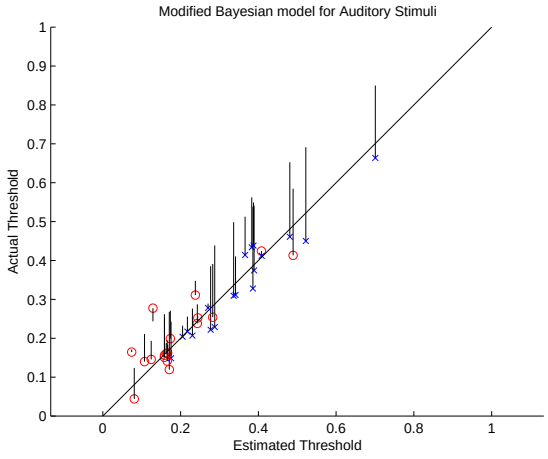


Figure 3.19: Modified prior results in slightly better estimates in congruent condition (not significantly better).

The results of these new weightings are illustrated in figure 3.19. This modified bayes rule predicts the actual values slightly (but not significantly) better than the unmodified bayes rule (Likelihood ratio:

		p	t
Modified Bayes	Low Coherence	0.44	0.79
	High Coherence	0.11	-1.67
Original Bayes (flat prior)	Low Coherence	0.06	2.02
	High Coherence	0.59	-0.55

Table 3.3: Frequentist t-test results for two Bayesian models, comparing the models to the actual congruent bimodal results.

Condition	Responses in direction of greatest coherence (%)	Weights	
		Visual	Auditory
Visual $>$ Auditory	71.8%	1	0.465 ± 0.005
Auditory $>$ Visual	59.6%	0.535 ± 0.005	1

Table 3.4: Weight changes between conditions

2.96, threshold: 9.87). Table 3.3 shows the equivalent frequentist results comparing the predicted values to the real values.

3.4.8 Conflict condition and the Normalisation Model

The Normalisation model has, as parameters, two weights, controlling the strength of the connection from the unimodal regions to the integration region. So far these have been kept equal, for the congruent condition, under the assumption that the two stimuli are equally weighted. In the conflict condition, the Normalisation model with tuning curve widths of 3 (and equal weights) will respond in the direction of the more coherent stimulus 53.2% ($\pm$0.45%, 95% CI) of the time. This is lower than either of the actual results (vis: 71.75%, aud: 59.63%). One might expect that the cause of this inaccuracy is that the two inputs are not integrating, due to the tuning curves being quite narrow. Widening them (even to many times this width) did not result in an improvement in the model's accuracy. It turns out that narrower tuning curves caused the average estimate to move towards the more coherent stimulus. This was due to a fall in the number of times integration occurred, leaving the more coherent stimulus causing the greatest peak.

To improve the fit to the data, the Normalisation model's two inputs were weighted differently. The more coherent modality had a weight fixed at one, while the least coherent modality was allowed to vary. Table 3.4 summarises the weights which fit the two conditions best.

The model can explain the results with only a slight change between the two conditions in the size of the connection weight, with the visual input weighted 15% more than the auditory weight. Putting this weight difference into the normalisation model for the congruent condition gives us a new set of

predictions, which are very similar to the earlier estimates provided by both the Bayesian model and the previous (equal-weight) normalisation model. The log ratio test between this model and the bayesian estimates has a log-likelihood difference of 2.99, the threshold is at 9.87 at the 5% level, so not a significant difference in the accuracy of the model results.

3.4.9 Adaptation Effect

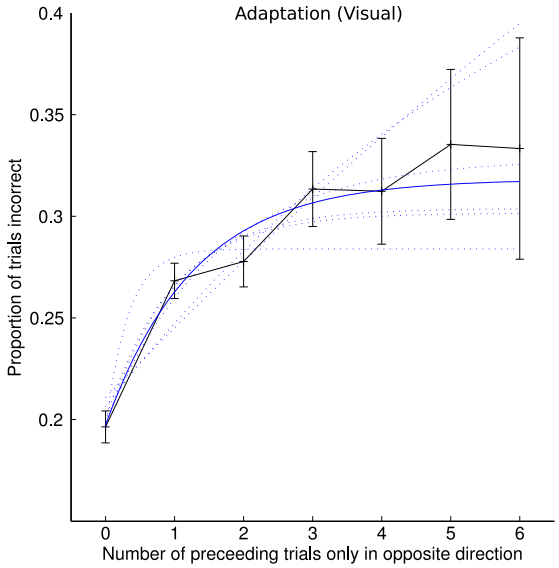


Figure 3.20: Effect of multiple trials of the same condition. Visual Stimulus. In this and the following figures, the first bar (zero) refers to all trials, (i.e. the previous trial will be in the same or opposite direction). The standard error confidence intervals were taken into account when fitting. The dotted lines indicate standard-error confidence intervals for the three parameters. The fitted curve was of the form $y = a + be^{cx}$. In this case a,b and c were: 0.3182, -0.1209 and -0.7801 respectively.

Table 3.5 shows the number of trials for each condition. As predicted, there are more mistakes in a trial following a trial of the same direction. Chi-squared tests found the three tables all significantly different from chance (all, $p < 1 \times 10^{-9}$; visual χ^2, 126; auditory χ^2, 39; bimodal χ^2, 76).

Effect of consecutive trials

Figure 3.20 illustrates how the adaptation effect mounts up with several trials of the same condition, for visual stimuli. An exponential decay function was used to model the effect of additional trials (it seemed reasonable to assume that additional trials would have proportionally less effect). For auditory and bimodal stimuli this effect doesn't seem as robust (Figures 3.21 and 3.22). To test if adaptation increases with more trials a linear regression fit was calculated for the different conditions. A linear fit was chosen to test for an increasing effect of adaptation because, although clearly the pattern is

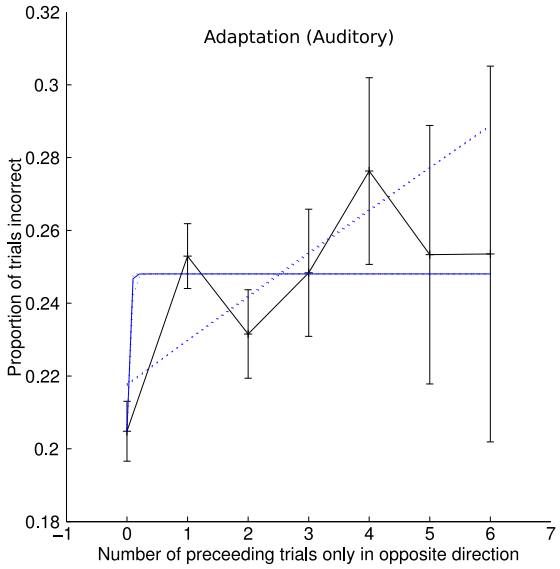


Figure 3.21: Effect of multiple trials of the same condition. Auditory Stimulus. See caption in Figure 3.20 for details.

not linear, the question of interest was whether further trials would induce further adaptation. The exponential fit was more ambiguous due to the extra parameter, which made inference about growth from the parameters determined through curve fitting difficult. In summary, the relatively large error bars meant that estimates using the sensitive exponential function were far less robust that estimates using the less-plausible but more reliable straight-line function. The first point (at zero) in the figures refers to the 'all trials' condition, (i.e. the previous trial will be in the same or opposite direction). This data point was not used in fitting the linear regression parameters. Table 3.6 summarises the gradients in the three conditions. These gradients refer to the proportion of trials extra we would expect the participant to get wrong for each extra trial they had seen of the same direction. The difference in gradient between the visual and bimodal conditions was significantly different (x = 1.86, p = 0.03, not significant when the two comparisons are taken into account) but was not significant between the visual and auditory conditions (x = 1.62, p = 0.052).

In summary, additional adaptation clearly occurs in the visual modality, but appears absent in the bimodal condition. The bimodal condition does have one fewer values (due to a slightly smaller sample size), but this slight difference almost certainly can not explain the difference between the visual and bimodal conditions.

Actual and Perceived Stimuli

Are participants anticipating and expecting the stimulus direction to reverse? One argument which suggests that this is not a conscious decision is that the adaptation effect varies between stimulus types (log-linear model for a $2 \times 2 \times 3$ contingency table, testing for block independence between the three

Sound Only	Correct	Incorrect
Same	1781	603
Different	1983	433

Vision Only	Correct	Incorrect
Same	1896	695
Different	2237	372

Bimodal	Correct	Incorrect
Same	1511	529
Different	1834	326

Table 3.5: Contingency tables for adaptation analysis.

Condition	Gradient (%/trial)	standard error (%/trial)	p
Visual	1.43	0.51	0.0025
Auditory	0.27	0.5	0.29
Bimodal	-0.04	0.60	0.53

Table 3.6: To test whether the effect of adaptation increases with more trials, without the effect of the first condition, the increase in errors per extra trial in same direction was calculated. p values not corrected for multiple comparisons.

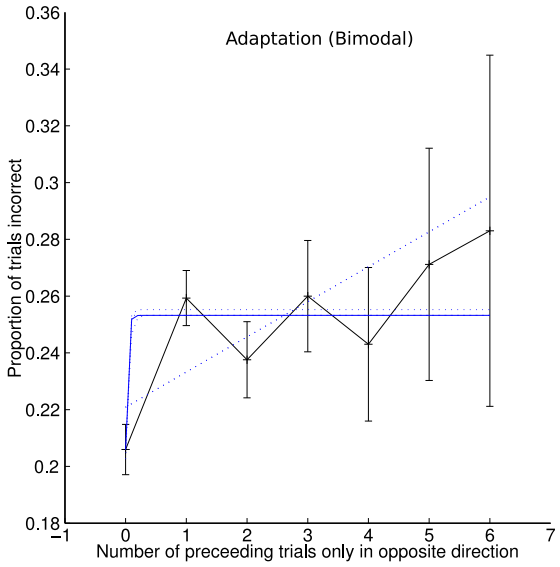


Figure 3.22: Effect of multiple trials of the same condition. Bimodal Stimulus. See caption in Figure 3.20 for details.

stimulus types. Deviance = 16.7, 3 degrees of freedom, χ^2 test, p = 0.00082), this suggests that the adaptation is not due to a conscious decision to swap directions regularly, as one would expect the same effect to be found across stimuli types if the influence is due to a conscious decision process.

The adaptation effect, so far, has focused on the effect of the actual stimuli. An alternative perspective is to consider the perceived stimuli. If the adaptation effect was due to changes in a high-level integration, perception or decision making region, then we would expect the adaptation effect to be stronger when considering the decisions the participants made (which we would expect to reflect their perceptions) than when using the actual stimulus directions. The adaptation effect was modelled, and fitted to the experimental results. Then, using this model, adaptation over the perceived directions was estimated (Table 3.7).

A χ^2 test on each pair of perceived-adaptation conditions finds the model in the sound-only condition has significantly less adaptation in the modelled perception-adaptation condition than the experimental results before multiple comparisons correction (auditory, visual, bimodal conditions; χ^2, 5.13, 2.52, 0.91; p-values (1 d.f), 0.02, 0.11, 0.34). None are significant after bonferroni multiple comparisons corrections. This suggests he adaptation is largely associated with he stimulus, not higher level perceptual or decision making areas.

Adaptation and the Normalisation Model

The Normalisation model was modified to model adaptation, with a parameter ρ describing the strength of adaptation. Figure 3.23 illustrate the effect of the adaptation modification on the model, by using an exagerated value of the ρ parameters. Figure 3.24 follows a similar sequence, but uses the fitted values

a) Sound Only

	Model		Experiment	
	Different	Same	Different	Same
Stimulus	1983	1781	1983	1781
Perceived	1939	1824	2009	1755

b) Visual Only

	Model		Experiment	
	Different	Same	Different	Same
Stimulus	2237	1896	2237	1896
Perceived	2190	1943	2241	1892

c) Bimodal

	Model		Experiment	
	Different	Same	Different	Same
Stimulus	1834	1511	1834	1511
Perceived	1768	1578	1795	1550

Table 3.7: The results of the adaptation model and the experimental adaptation data. Each table is for a different stimulus type. The top row of each table are the results when using the stimulus to determine whether the previous trial was the same as the current trial. The lower row of each table uses the responses of the real (or modelled) participants to determine whether a trial was in the same or opposite direction. The first two data columns are the model's results and the second two are the experimental results. Note that the model's parameters were fitted so the stimulus adaptation matched. In summary the comparison of interest is between the perceived model results and the perceived experimental results.

Region	ρ
Visual Unimodal area	1.07 ± 0.01
Auditory Unimodal area	0.49 ± 0.01

Table 3.8: Adaptation constants in the two regions, for unimodal stimuli.

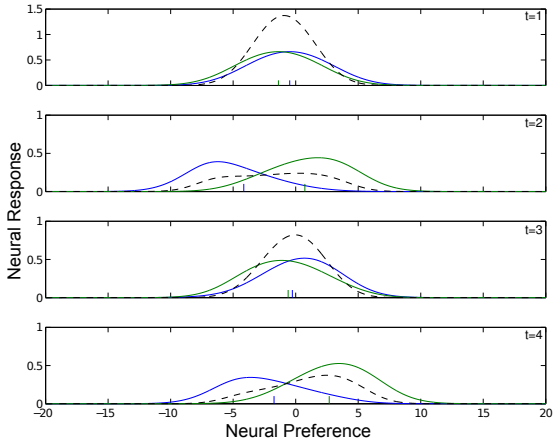


Figure 3.23: Normalisation Model: Adaptation using exaggerated ρ value to illustrate its effect on the model. The top plot illustrates the neural response in the first trial, the second plot the response in the second trial, etc. Coloured vertical tick marks on the x-axis illustrate where the peaks of the inputs would lie if the adaptation effect was removed. In the second trial the response would have been to the left of zero, but the peak of the bimodal distribution is on the right due to the effect of the previous trial. Blue curve, visual neural responses; green curve, auditory neural responses; black-dashed, bimodal neural responses.

of ρ (Table 3.8).

Table 3.9 indicates the adaptation effect of the unimodal-only adaptation-modification model is close to that of the real bimodal adaptation result. This similarity suggests an adaptation effect in unimodal regions continues to influence the integration region's results. Oddly both in the actual and modelled results the bimodal adaptation does not lie evenly between the two unimodal conditions but appears to remain approximately as high as the highly-adapting visual condition.

The model was altered to incorporate adaptation at both the unimodal and bimodal stages (also in table 3.9 and in figure 3.25). The adaptation effect in the auditory region was set to zero, to find the range of possible values the bimodal region could adopt. It was found in the auditory condition the bimodal region needed to have an adaptation parameter of 0.86 for the auditory adaptation to match the experimental results. Similarly, fitting the visual condition found that the visual layer required a ρ of 0.42. With these settings for adaptation it was found that the bimodal condition experienced an adaptation effect of 19.5%. 56% higher than the experimental results. It seems that transferring the adaptation mechanism to the bimodal region resulted in an overall increase in adaptation in the bimodal effect, above the adaptation recorded in the experiment. It was found that an adaptation effect of even just $\rho = 0.2$ in the bimodal region resulted in an over-estimate of the bimodal adaptation (once fitted to the unimodal adaptation results). We tentatively can conclude that a large majority of the adaptation must be restricted to the unimodal regions.

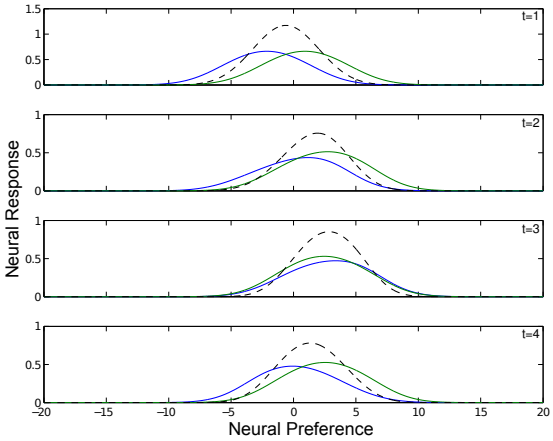


Figure 3.24: Normalisation Model: Adaptation using estimated parameter values. The top plot illustrates the neural response in the first trial, the second plot the response in the second trial, etc. The effect is more subtle. It is most clear in the shape of the blue curve in the second plot, which is skewed towards positive values due to its earlier response peaking at a negative value. Blue curve, visual neural responses; green curve, auditory neural responses; black-dashed, bimodal neural responses.

Coherence and Adaptation

Our final hypothesis investigates whether adaptation occurs more in the low or high coherence trials. Although not explicitly modelled we would predict that the adaptation effect would be stronger in the low-coherence condition for the simple reason that with more errors occurring there would be more opportunity for adaptation to take place; as the number correct approaches 100% the difference in the number correct between the same and different conditions will approach zero. In the analysis so far we have treated the two conditions identically for the purposes of adaptation. Although the trials analysed were interleaved.

The low coherence condition has considerably more adaptation than the high-coherence condition (Table 3.10 illustrates this difference. Because there would be differences in the number of correct trials between conditions, the log-linear model could not test this difference. Instead, a contingency table of just the correct trials was used. This simply tests for if there was a significant effect of coherence on the same/different condition results. The results were, as expected, a significant difference ($\chi^2 = 5.99$, p = 0.014), with the low coherence having a greater difference in the number of correct trials.

Region	Actual Adaptation	Model (unimodal only)	Model (full-bimodal-region)	Model (mix)
	Adaptation Effect			
Visual	12.8%	12.7%	12.8%	12.6%
Auditory	7.8%	7.9%	7.7%	8.0%
Bimodal	12.5%	11.5%	19.5%	15.3%
	Adaptation Parameter ρ			
Visual		1.07	0.42	0.89
Auditory		0.49	0	0.34
Bimodal		0	0.86	0.2

Table 3.9: Size of Adaptation in each modality. Note that the Visual and Auditory models were fitted to the real data. The interesting result here is the similarity of the bimodal adaptation score to the actual adaptation. The top three rows indicate the results of the experiment (Data 95% CI: $\pm$ 4.9% across subjects) and the three model runs (Model 95% CI: $\pm$ 0.3%). The bottom three rows record the adaptation parameters in the three regions.

High Coherence	Correct	Incorrect
Same	1063	227
Different	1169	141

Low Coherence	Correct	Incorrect
Same	833	468
Different	1068	231

Table 3.10: Contingency tables showing effect of coherence on adaptation for visual unimodal stimulus.

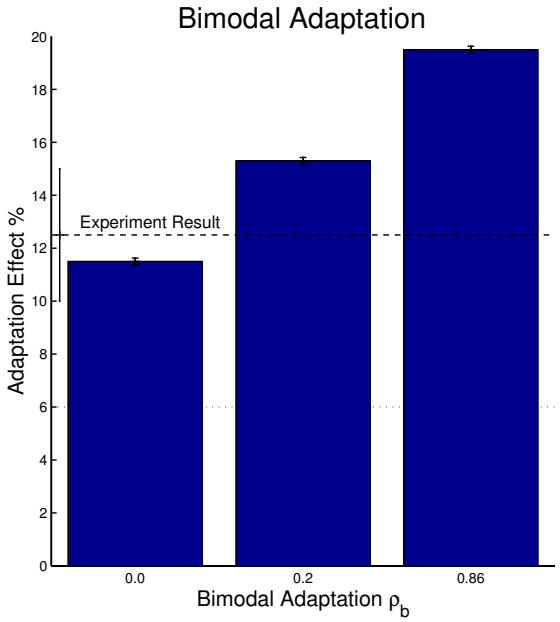


Figure 3.25: Plot summarising the results in table 3.9. Error-bars indicate one standard error of the mean. The experimental result error-bar computed across subjects. The three bars are for the three model-runs at the three different bimodal adaptation parameter values (0, 0.2 and 0.86). The unimodal adaptation parameters were fitted to the experimental results.

3.5 Discussion

3.5.1 Success of stimuli choices

The visual and auditory stimuli both achieved the behavioural requirements. First there was considerable stability between sessions, as required. Second, the two thresholds were successfully determined in both modalities with sufficient accuracy. Third, the congruent bimodal stimuli was well perceived with integration unexpectedly close to the Bayesian integration prediction. The only concern which presented itself in the results was in the conflict condition, with some subjects reporting some of the trials appeared to be in conflict. The intention had been for the conflict to be imperceptible.

A more general concern and one that has not been tested in this study, is whether it is the sense of a 3d trajectory which is being integrated here or rather the integration of simple low-level properties. For example, it might be a particular detail of an auditory cue (e.g. whether sound intensity from a particular cue increases on the left) combined with a simple visual cue (simple planar dot-motion), which provides the subjects with sufficient information that they are able to complete the task. This could be tested by training the participants on conflict stimuli from the start, and see whether similar levels of accuracy are attained. Alternatively, the axis of the visual motion could be rotated 90 degrees, to make the two motions orthogonal.

3.5.2 Model results

Bayesian

The accuracy of the Bayesian prediction was not completely a foregone conclusion. Recent discussion has suggested that humans have a prior preference to rely on visual stimuli over other modalities even when the coherence of the two modalities is equal (Battaglia et al., 2003). It was somewhat surprising then that the results followed the Bayesian predictions so closely. It appeared, at least initially, that this was a very good model. However, the lower coherence estimates were not as accurate as the higher coherence ones. Such deviations from the predicted bayesian optimal solution were then pursued as an opportunity to distinguish the competing models of neural integration.

The conflict condition was very different, with the Bayesian model performing very poorly in the low-visual coherence condition (while predicting well the high visual coherence condition). This failure in the conflict condition indicated the basic Bayesian model was not sufficient. It is important to note that the conflict condition was perceived by many of the participants, making results from that session somewhat difficult to interpret.

Normalisation

As an alternative to Bayesian integration we considered the model described in the recent paper by Ohshiro et al. (2011). In this model, two layers of unimodal neurons model the primary sensory regions. The activity of these neurons is combined in a third, integration, region. The signal from each of these

bimodal neurons is divided by the sum of all the other output neurons, which causes a form of divisive normalisation. Combined with plausible non-linearities, this system produces a response similar to the Bayesian integration of two probability distributions. The output is only approximately Bayesian, and differences may exist which would allow the two models (Bayesian and the equivalent divisive normalisation-implementation) to be distinguished.

To apply the Normalisation model to the psychophysics results, several assumptions were applied regarding the behaviour of the input sensory neurons. In particular they were assumed to be non-noisy, instead responding deterministically to a randomly positioned stimulus. Future studies could modify the model to make the neurons respond in a stochastic manner.

In the Bayesian Model, it would not matter how far apart the two stimuli were, as the Gaussian distributions representing the two modalities remained non-zero for an arbitrary distance from the mean. The use of limited receptive fields in the normalisation model meant that when the estimates in the two modalities were sufficiently distant the two stimuli were not integrated, causing a reduction in the accuracy of prediction below that of the Bayesian integration model. This distinction may reflect the Bayesian model's failure to predict the results of the conflict condition.

There was a lower accuracy in the lower coherence stimuli, but this did not significantly distinguish the two models. The Normalisation model fitted the data well, with a possible trend towards being better than the Bayesian model. To investigate further, the Normalisation model was modified to fit the model to the conflict condition results to find the weight each modality should be given. Intriguingly only a slight difference was required to generate the bias towards the visual stimulus reported in the conflict condition. This sensitivity in the conflict condition to differences in the weighting of the two modalities might make it a useful stimulus in future experiments. Even with the modified weights, the model was still successful when integrating the congruent stimuli. It seems the normalisation model may be relatively immune to small weight differences when the two stimuli are congruent.

Modified Bayes

The Bayesian model was also modified to include a prior preference for the visual modality. This corresponds to Cheng et al. (2007)'s second type of integration, in which previous knowledge about stimuli is integrated with the current stimulus. Battaglia et al. (2003) reported that both the maximum likelihood estimation and the visual capture model were partially correct. In their study, subjects combined both visual and auditory stimuli but favoured the visual over the auditory. To interpret the results they also introduce a prior which increases the model's reliance on visual information. In our study the parameters for a prior distribution were fitted using the conflict condition (as mentioned above, potentially contaminated by conflict-detection), and then reapplied to the congruent condition. Intriguingly the normalisation model's reduced accuracies in the low-coherence condition were mirrored in the modified Bayes model too. In future experiments, additional conflict condition trials should be conducted to test this hypothesis more thoroughly. For example, the hypothesis implies the results should all be predicted using a prior distribution with a single fixed parameter A. Further experiments could ask whether this

parameter remains constant for a variety of different visual/auditory integration tasks. For example, is the value of *A* reported here only valid for self-motion cue integration?

Visual Capture

Although, in general, it appeared the favoured models were those close to Bayesian predictions, the conflict condition sessions were found to be very non-Bayesian (at least using our naive Bayesian integration model). It appeared many of the participants had limited or no integration of the two conditions. There is a reasonable likelihood that this is because of partial detection by the participants. Many of the participants in the high-visual coherence condition appeared to select the visual stimulus and ignore the auditory (indicated by the cluster of results around the 84% accuracy level). This suggests the visual capture model may often occur. The equivalent condition (with auditory stimuli holding a higher coherence than the visual) does not have the same cluster around the 16% threshold. This suggests this is an effect particular to the visual modality, as predicted previously (Alais & Burr, 2004).

The mechanisms which decide whether two stimuli are associated with the same cause may be at the heart of the conflict condition's results. Our interpretation of the results possibly should include a higher level of processing suggested by Roach et al. (2006), which determines whether the cues should be integrated, although in some ways the Normalisation model already does this by the limited tuning curve width and non-linear super-additivity. As above, a new experiment is needed with varying levels of congruence and conflict in the two stimuli. Such an experiment should incorporate the ideas of causal inference, as described by **?**, in which an ideal observer is assumed to model the likelihood of two events having the same cause. In general it seems including an interaction prior allows more accuracte modelling of psychophysics results for a wide variety of intergration tasks.

3.5.3 Adaptation

The strength of the adaptation effect was quite surprising, given that the experiment was not initially designed to find such an effect. In all three stimulus conditions the adaptation effect was highly significant, with the intriguing, although possibly inevitable finding (given the size of the main effect) that the adaptation effect varied between conditions, with visual and bimodal conditions having greater adaptation.

In retrospect it would have be advantageous to the experiment's initial design goals to order the sequence of trials types to minimise adaptation, as it can (and will have) affected the results somewhat. In particular in the higher adaptation conditions it will have lead to an increase in the coherence thresholds to compensate for the poorer performance.

Additional trials

Additional trials of the same direction caused further adaptation in the visual modality, but this was not significant in the auditory or bimodal conditions. One hypothesis for the weaker integration of the effect

over trials for the auditory condition could be the transitory nature of the auditory stimulus. The ‘sense of motion’ may only exist for a limited period during the trial, and therefore neurons responding to this stimulus have more time to become re-adapted. In hindsight, this may make this stimulus a poor choice for the following fMRI study.

Adaptation Model

The Normalisation model was altered to include a simple adaptation effect by making neurons have a reduced firing rate after a trial in which they experienced high activity. Each region had one parameter describing the adaptation effect. These were fitted using the unimodal sessions. The model was then compared to the experimental results in the bimodal condition. Although no statistical tests were performed, both the model and the experimental results appeared to indicate that the adaptation remained at approximately the high level of the adaptation experienced in the visual unimodal condition. Importantly, when the adaptation mechanism was introduced to the integrating layer, the bimodal model results increased well beyond the values found experimentally. The agreement of the model and the experimental results in the unimodal-only adaptation suggests that adaptation does occur at the level of the unimodal inputs. A difference, for instance a marked underestimate of the adaptation, would have suggested additional adaptation must occur downstream.

Perception and Adaptation

The weakness of the effects of adaptation in the perceived direction results is quite intriguing. At its most interesting it may point to the adaptation effect being associated with regions prior to both integration and decision making. One might argue that if it were otherwise, and the integrated, decision representing neurons adapt, the adaptation effect would be reflected more strongly in the results of the perceived directions of motion. However, one important caveat regarding these conclusions is to consider whether this type of adaptation (linked with the perceived stimuli) could have been masked by the stimulus-associated adaptation. In particular, in those trials in which adaptation occurred in a low-level cortical area, adaptation in the same direction would have occurred at the cross-modal or decision-making layer. Put simply, adaptation will occur in a similar direction in all areas of interest, which means the effect of one over another will be difficult to discern.

Still, I was expecting a larger effect of the perceived-response adaptation than that found. How do we reconcile this apparently quite weak result with the findings of electrophysics experiments which suggest VIP neurons do adapt (Bremmer, 2005, and figure 4.5)? Finding little adaptation in integration regions also seems to contradict studies which have found behavioural adaptation in more abstract stimuli such as facial expression (Adams et al., 2010), and findings by Boynton & Finney (2003) suggesting that adaptation to orientation becomes more pronounced (for adapting stimuli of a few seconds) as one moves from lower to higher visual areas. One possibility is that much of the adaptation at higher levels is inherited (Krekelberg et al., 2006). However, adaptation often appears invariant of features considered independent at low-levels (such as size). Further, some studies, such as Ehrenstein & Reinhardt-rutland

(1996), have reported cross-modal adaptation between visual and auditory motion cues, although in their study the visual adaptation stimulus lasted 150 seconds (twenty times longer than in this study).

In summary, although a potentially interesting technique, which may allow the low- and high- level effects of adaptation to be separated, the results are equivocal regarding the localisation of adaptation to either high or low level processing. Features which are invariant in particular processing domains (including those which can be presented in a cross-modal fashion) make much better stimuli for determining the extent of adaptation at different layers of processing. However, determining the level of adaptation within decision making processes may be more difficult to deduce and a method similar to the simple model used here may allow such features of perceptual processing to be detected and quantified.

Adaptation in the Bimodal condition

It was found that the low coherence condition has considerably more adaptation than the high-coherence condition. This is to be expected as the low coherence condition has more errors and lower signal and so more opportunity for the wrong choices to be made.

Could this extra adaptation cause the sub-optimal accuracy in the bimodal low-coherence condition? One would expect that the adaptation effect (which could cause such a drop in accuracy) to also exist in the two unimodal conditions. Therefore the reduction would already be taken into account in the estimates of the bimodal thresholds. However, the bimodal condition seems to have the high-adaptation of the visual unimodal condition, even though one might imagine it should reach an average of the two. A naive integration model would assume such errors would be independent and therefore reduced or averaged when integrating the two unimodal stimuli. Clearly the experimental and modelled results both show this reduction does not take place. It is plausible then that the slight non-Bayesian integration result is due to unexpectedly large levels of adaptation, and not other hypotheses suggested earlier. A final caveat or hypothesis; We know the stimuli coherences in the bimodal condition were lower than in the unimodal conditions (to reach the same threshold). These lower coherences might explain the unexpectedly large adaptation in the bimodal condition: The sensory areas will have had a weaker signal allowing adaptation to act more on the responses, which is reflected in the bimodal result. It is interesting, if this is the cause, that it is reflected in the Normalisation model's predictions.

A final possible explanation for the values of adaptation in the different conditions is the influence of a contingent aftereffect. This is where one feature of a stimulus can become quickly associated with another feature. This has been found to occur in both visual and auditory modalities (Dong et al., 1999, 2000). The bimodal condition was congruent until the last session (including during training). It is during these sessions that the two cue types would become associated, and therefore the presentation of one would induce the perception of the other. This does raise the concern that such integration is not self-motion specific but may be due to a highly generalised ability to associate disparate cues (from different modalities). A way to test this is to perform the same experiment but with other cues not intended to represent self-motion and see whether similar integration can occur. For example Calabro et al. (2011) used spatially co-localised cues and controlled for such concerns by using similar spatially

remote auditory cues. The investigation of general-integration proficiency is beyond the scope of this thesis.

3.5.4 Implications for connectivity

The apparent reweighting of the visual and auditory stimuli in both the Bayes and the Normalisation model (for the conflict condition) suggests the dynamic modulation of connections occurs between the sensory and integration regions. If the connections remain fixed, the reliability of the two inputs must be incorporated in another way, outside the current model. One problem with the model is the assumption that the sensory population has a Gaussian response to the stimuli. The likely response is far more noisy, with the noise potentially providing the reweighting currently provided by the modulation of the weights from the unimodal regions, as in the model by Ma et al. (2006).

Whether the integration is complete or due to cross-modal interaction is difficult to determine. The adaptation results find the actual directions cause more adaptation than the perceived directions, suggesting adaptation happens in neurons in sensory areas. The distinction in the strength and type of adaptation in the different modalities also suggests low-level cross-modal interactions are limited. If cross modal interactions were major drivers of integration one would expect the adaptation to be roughly equal across modalities, as the same populations of neurons would respond.

3.5.5 Future Work

Psychophysics

To distinguish between models it is clear that future experiments need to concentrate on the responses to stimuli at very low coherences. I would suggest several thresholds between 60% and 75% accuracy should be determined for integrating stimuli. It may also be advisable to simplify the stimuli, and use planar motion and rotating audio cues. This would make the stimuli much easier to create and interpret.

General

The large variance in the results across subjects for the conflict-condition may have been to do with different strategies or attentional differences between the subjects. However, the spread of responses does offer a potential method for investigating differences between subjects in how they integrate stimuli. Initially I attributed the variation in the subject responses to differences in strategy or attention. Intriguingly the Normalisation model and the modified Bayesian model both showed that the proportion being reported in the direction of the most coherent stimuli could be adjusted with only small changes to a single parameter, with very little effect on the results in the congruent conditions. If the differences in subject responses was due to slight differences in such a parameter (and not due to higher-level strategy differences) then such sensitivity may open the door to future experiments, in particular those associated with changes in cognitive function, such as those linked with ageing or dementia.

Adaptation

The modified Normalisation model allows precise predictions regarding the behavioural response during adaptation. An experiment could be constructed which carefully causes adaptation at different locations in the space of parameters being estimated. The tuning curves of input neurons are believed to be modulated by coherence or attention. By combining such changes with a cross-modal adaptation experiment, the widths and interactions both within regions and across the hierarchy could be refined more accurately. For example, the hypothesised narrowing of tuning curves due to attention would become apparent by the restriction in the level of adaptation to stimuli close to the adapted direction. Similarly, investigating the level at which adaptation occurs can be done by looking for adaptation across modalities. Note however that the model does lack the top-down connections which might cause adaptation in sensory areas unassociated with the stimulus; simultaneous refinements will be required within the model while applying it to new experiments.

Probability Summation

Other models of integration exist, which have not all been investigated in this thesis. For example Wuerger et al. (2003) investigated the detection of a moving stimulus and suggest its detection is not through the integration of the two stimuli but rather occurs at a 'higher' level of *probability summation* which is blind to the direction of motion, and purely combines the likelihoods of motion in the two modalities. This study looked at distinguishing motion, rather than detecting motion, so was not well suited to expanding on the cited study. Adaptation may assist in this case too, if cross-modal behavioural adaptation is shown to exist, it would suggest integration is of stimulus estimates and not probabilities of particular features.

3.5.6 Summary

Although initially intended as an experiment to perfect the stimuli for the fMRI study, the results reveal several interesting features of perception and integration and aided the development of several alternative models for integration.

With regards to its initial purpose the experiment was successful, with most participants able to perform the task, the results were stable and integration was found between modalities. Of the models used, the standard Bayesian model closely predicting the results of the experiment. The naïve alternative (no integration) was clearly invalid. The modification of the Normalisation model proposed by Ohshiro et al. (2011) for use with a psychophysics experiment allows it to be tested indirectly through behavioural results. There was a trend towards slightly under-performing compared to Bayesian integration in the experimental results, so the Normalisation model may be a better estimate than the standard Bayesian integration. However, the incorporation of priors to weight particular modalities appeared capable of predicting the experiment's results in a similar direction as the Normalisation model. It seems possible that further psychophysical experiments could be developed which could distinguish models,

if developed with that task in mind.

The adaptation effect was much larger than expected, and had interesting differences between modalities. In particular, the bimodal condition's adaptation, when compared to the unimodal conditions appeared to match the higher of the two unimodal conditions. This was reflected in the modified Normalisation model. The importance of adaptation in the integration results is not clear however, but adaptation may be an interesting avenue to further investigate integration at low coherence. The modified Normalisation model could include adaptation in both the unimodal layers and the integration layer. However, it was found that adaptation in the integration layer would cause the bimodal results to have excessive adaptation, beyond that recorded in the experiment. So it was concluded that most of the adaptation appears to be in the unimodal regions.

Implications for the fMRI study

The experiment was initially conceived as a trial-run of the stimuli in preparation for the fMRI experiment to follow. The first aim was to confirm that participants were able to use and understand the stimuli to complete the task. Second, the range of possible coherences the two stimuli needed to cover was explored and during pilot experiments the stimuli were designed to minimise floor or ceiling effects. Third we wanted to confirm that the two thresholds (at 71% and 84%) were discernible within the experimental time window available. Finally we needed to test whether the two cues were being integrated or if participants generally ignored one and used the other. Clearly all these targets were met, and so confirmed the stimulus was ready for use in the main fMRI experiment.

Beyond validating the stimulus, the experiment also suggests that changing the coherence of the two stimuli appears to strongly influence which is used. Such a process could be mediated by connectivity changes between the unimodal and integration regions. Second, the adaptation effect described above is likely to be due to neural adaptation. This is a feature which can be detected and exploited in fMRI experiments. For example cross-modal adaptation can provide a clear indication that a neuron is bimodal. The Normalisation model predicts superadditivity, although other methods exist which do not, such as the model proposed by Ma et al. (2006). The following fMRI study may help distinguish these two models through the presence of changes in connectivity and superadditivity, both features predicted by the Normalisation model and both features absent in the Ma et al. (2006) model. Adaptation may vary between regions in a way which follows the predictions in this chapter, with the unisensory regions (the visual and auditory cortices) experiencing much greater adaptation than the putative integration region (the ventral intraparietal sulcus).

Chapter 4

Self-motion: An fMRI study

The previous chapter provided strong evidence for the successful integration of self-motion cues, with hints through the use of the normalisation model that the integration may occur in a separate integration area, rather than through cross-modal interactions between the two sensory regions. Whether this theoretical hierarchy is valid is almost certainly only testable by resorting to direct analysis of brain activity. In this study we used fMRI to test whether a region previously identified as a strong candidate for integration (a possible human analogue of the primate ventral intraparietal area) is responsible for integrating the two signals. We also investigate the weighting of the connections from unimodal areas to the putative integration region. We hypothesise that the region associated with the high coherence modality will have a stronger connection to the integration region. Other features also examined by this study include cross-modal adaptation to test our hypothesis that the same neurons are responding to congruent directions between the two modalities. Pattern classification is used to test our hypothesis that the direction of motion can be decoded from regions such as the human putative VIP area.

4.1 Introduction

4.1.1 Study overview

Recently, considerable debate has surrounded the ventral intraparietal area (VIP) and the possibility of a human equivalent region in the intraparietal sulcus (IPS) as a self-motion integration region (section 4.2.4), but there is an absence of human brain imaging studies looking at this potential purpose using either realistic visual or auditory motion stimuli. To try to bridge these gaps in the research field, this study will introduce new audiovisual stimuli and investigate audiovisual integration in putative human VIP. The fact that a region in the IPS of humans responds to both visual and auditory motion is not sufficient proof that the region integrates the stimuli. To test whether this region does integrate these cues, one needs to know if the same neurons are responding to the two modalities. More specifically for

this study, whether the same neurons respond to the same direction of motion, regardless of the stimulus modality.

4.2 Background

4.2.1 Visual Motion Regions

Overview

Visual stimuli are first detected and processed in the retina, The signals are then passed via the lateral geniculate nucleus in the thalamus to the primary visual cortex (V1), also known as the striate cortex. Further processing is distributed among a hierarchy of cortical areas across much of the parietal and temporal lobes (Orban, 2008). This thesis is concerned with motion (specifically that which is consistent with ego-motion). The review by Britten (2008) summarises the regions which are likely to be involved in self-motion visual processing and perception. Figure 4.1, from his review, illustrates how the hierarchy of visual motion regions is connected. Of particular interest are the human equivalent to primate regions MT, MST and VIP. For a more detailed assessment of visual self-motion processing, refer to the review and for a review of multisensory integration (focused on motion) see Beauchamp (2005) or Fetsch et al. (2010).

Note that since Britten's review, another region has been reported to be important for visual motion. Wall & Smith (2008) found an area, dubbed the cingulate sulcus visual area (CSv) which appears to be selectively responsive to visual selfmotion (Cardin & Smith, 2010). Which primate area this is homologous to is still uncertain, or the region's purpose or response characteristics.

It is also worth remembering that the processing of optic flow cues into navigation relevant information occurs throughout the sensory system and beyond. For example, a study looking at the effect of attending to one's change in orientation found that a series of locations had increased activity during the attention periods (Diekmann et al., 2009). Attention appeared to modulate activity in regions such as the hippocampus and several parietal areas, although none of the significant clusters were reported to be in the vicinity of the putative location of the human equivalent of VIP.

V1, MT and MST

In non-human primates a region known as the middle temporal (MT or V5) is found within the superior temporal sulcus and appears to be principally involved in visual motion perception. A region referred to as the human MT complex (hMT+) also responds to visual motion stimuli and, using retinotopic mapping fMRI, has been found to consist of at least two subfields (Huk et al., 2002). It is assumed that this region is in some way homologous to regions MT and MST found in other primates, due to its location and particular motion-selectivity. Some of the methods used to distinguish these two subfields rely and are based on previously reported differences in their responses to various stimuli during electrophical recordings. Examples include the difference in single cell recordings during pursuit eye

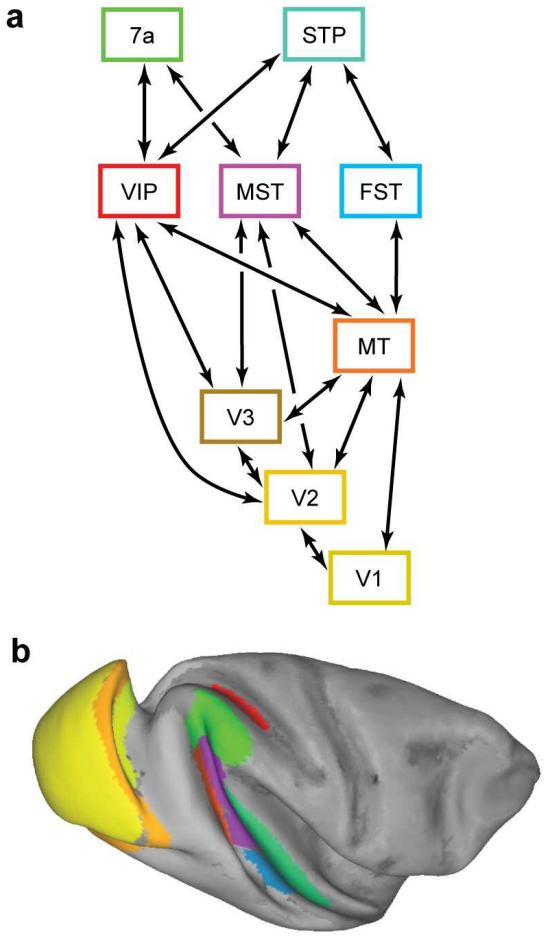


Figure 4.1: A schematic of the anatomy of the visual motion system, from Britten (2008), figure 5. The anatomy is for monkey, with a slightly inflated monkey brain on which the regions are indicated. Note the apparent hierarchy, from 'early' visual regions (such as V1) to more specialised response areas such as MT and VIP.

movements. Newsome et al. (1988) found MST neurons had a higher firing rate during smooth pursuit, while MT neurons did not. This differences was exploited by Dukelow et al. (2001) to distinguish the two subfields using fMRI. Similar differences include the extent of receptive fields (in MST the larger fields have been found to extend into the ipsilateral hemifield), the coherence of MT's retinotopic map, compared to MST. Finally, both electrophysiological recordings in monkey (Duffy & Wurtz, 1991) and fMRI imaging in humans (Morrone et al., 2000) have found that MST and possibly an equivalent within human hMT+, responds selectively to complex visual motion (including circular and radial motion). The orientation of the two regions also provides some credibility for this homologue, with MST more anterior to MT in both macaque and the putative human equivalents found using fMRI. It is worth noting before moving on that hMT+ is likely to contain more than just homologues of just MT and MST.

For example the fundus of the STS (FST) appears to be a distinct region in some primates (Rosa et al., 1993), and region MST is generally considered to consist of at least one further division into lateral and dorsal portions. This means any localiser defining MT and MST will almost certainly have included other regions in those ROIs.

Both non-human and (putative) human V1 and MT have relatively small receptive fields (Albright, 1984; Engel et al., 1997, electrophysiological recording in macaque and a human fMRI study, respectively) and respond to most types of visual motion stimuli (Weliky et al., 1996; Xu et al., 2004; Kaskan et al., 2010, optical imaging in ferret, prosimians, monkeys, respectively), but don't appear to be specialised for self-motion visual stimuli (Wall & Smith, 2008, human fMRI study). The neighbouring region, MST (probably consisting of two functionally distinct areas), appears to be far more responsive to self-motion stimuli (e.g. Bremmer et al., 2010; Gu et al., 2006, electrophysiological studies in monkies). MST, besides its importance in the visual-motion pathway, has sizeable responses to vestibular cues (Fetsch et al., 2007, electrophysiological recordings in monkey) and pursuit eye-movements (Newsome et al., 1988, electrophysiological recordings in monkey), with eye-movements also being compensated for (e.g. Shenoy et al., 2002, electrophysiological recordings in monkey). fMRI studies suggest that MST, although multimodal (receiving visual, eye-movement and vestibular information), does not appear to respond to auditory motion cues in sighted adults (Bedny et al., 2010; Lewis et al., 2000).

This study will ideally use hMT (the human homologue of MT) as the region representing the unimodal visual signal. However, if the localiser is unable to distinguish hMT+'s subfields, the above summary suggests that the use of the whole hMT+ complex will not be corrupted by auditory stimuli.

Chapter 2 has more detail about MT's responses and activity.

Self Motion

The earlier areas (V1, MT, V3, etc) appear to have small receptive fields, and respond to simple linear motion. It is regions MST and VIP that interest us most, due to their large receptive fields and apparent preference for self-motion like cues. Of particular interest is the finding that VIP neurons appear to respond selectively to the location of the centre of expansion (Zhang & Britten, 2004). The cells recorded fell into two groups; sigmoid and Gaussian. In particular, most cells recorded had a maximum response (or gradient, for sigmoid cells) to self motion with an almost dead-ahead direction. The stimulus developed below tries to capitalise on this by restricting the centre of expansion to a point only 2.46° degrees from directly infront. A similar set of results, regarding a preference for self-motion-cues is found in MST (Page & Duffy, 2003, electrophysiological study in monkey), with cells also responding to congruent vestibular self-motion cues. It seems almost certain that both these regions are responsible for processing visual self-motion stimuli, and both appear to have some aspects of bi- or multimodal integration. It is the aim of this study to extend these findings from animals to humans using fMRI.

4.2.2 Auditory Motion regions

Input

Auditory stimuli are detected and transformed into electrical signals which are transmitted via the cochlear nerve to the cochlear nucleus. The information passes through three other subcortical regions before entering the cortex through connections to the primary auditory cortex (A1). fMRI and electrophysiology has found that A1 responds to most stimuli (including simple tones). The surrounding areas (often known as 'belt-areas' of the superior temporal gyrus) have greater selectivity (figure 4.2).

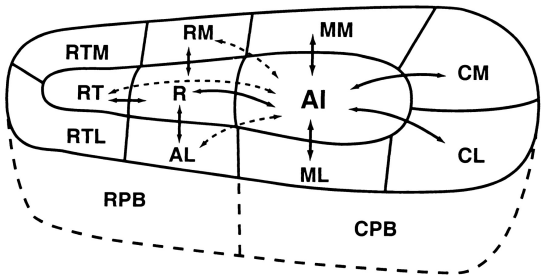


Figure 4.2: Figure from Kaas & Hackett (2000), figure 5 illustrating the connections within the Auditory Belt areas of the STG, from A1. In summary it shows dense connections from A1 to the surrounding core and belt (solid arrows). The connections with non-adjacent regions are less dense (dashed arrows). Importantly Kaas & Hackett (2000) note that the core has few, if any, connections with the parabelt or more distant cortex. Abbreviations of cortical areas, taken verbatim from paper: AI, auditory area I; R, rostral area; RT; rostrotemporal area; CL, caudolateral area; CM, caudomedial area; ML, middle lateral area; RM, rostromedial area; AL, anterolateral area; RTL, lateral rostrotemporal area; RTM, medial rostrotemporal area; CPB, caudal parabelt; RPB, rostral parabelt

What and Where streams

In contrast to visual motion processing, far less is known about the brain's processing of auditory motion. The processing of auditory stimuli has been compared to visual processing, with the suggestion that auditory processing is also divided into 'what' and 'where' streams. Evidence for such a division (in non-human primates) was reviewed by Romanski (2007). To summarise, the 'what' stream begins in the more lateral portion of the belt-region in the superior temporal gyrus and continues into the prefrontal cortex (Rauschecker & Tian, 2000). This organisation is reflected in the location of Broca's area, a cortical region associated with language.

The pathway of interest in this study, is the 'where' stream. Rauschecker & Tian (2000) suggested that this begins within the posterior portion of the superior temporal gyrus before being integrated with other stimuli in the posterior parietal cortex. Sestieri et al. (2006) found further evidence for this division, with ROIs specified in the fundus and wall of the IPS being associated with spatial processing, and regions in the superior temporal sulcus associated with recognition.

Does the brain process Auditory Motion?

Many papers have suggested auditory motion is a perceived feature of our environment and that particular regions of the brain respond to auditory motion, as opposed to simply increase activity due to greater calculation of purely stationary auditory directions (e.g. Alink et al., 2011).

The fMRI contrasts depend on the selection of auditory motion and control stimuli. Auditory motion stimuli have been generated by either changing the interaural time difference (ITD) (e.g. Smith et al., 2007), varying the interaural intensity difference (IID) (e.g. Lewis et al., 2000), differing frequencies between ears generating binaural beats (Perrott & Musicant, 1977) or convolving the sound with a head-related transfer function (HRTF) (Alink et al., 2011; Baumann & Greenlee, 2007). The stationary contrasting stimulus has usually consisted of the same stimulus presented from a fixed location or locations. The stimuli were almost always designed to be perceived as single sound sources, and the motion intended to reflect the object's motion, and not the motion of the participant.

Smith et al. (2007) criticise this choice of stimuli, suggesting the increase in activity could be explained by a non-motion sensitive brain region which requires additional processing to track the location of the auditory cue (also in their earlier paper, Smith et al., 2004). The auditory processing performed being agnostic regarding motion but will still have greater activity if required to update an apparent stimulus' location percept more regularly. It is not clear if this criticism applies to decoding experiments (such as Wolbers et al., 2011; Zvyagintsev et al., 2009).

In response Alink et al. (2011), performed a study which attempted to address Smith et al. (2007)'s concerns. They were able to decode the direction of auditory motion in the planum temporale, and possibly in A1 ($p < 0.10$). Note that the searchlight mean was subtracted from each voxel's time-course to try to mask univariate effects due to changes in the IID. Whether this is sufficient is unclear.

It is generally accepted that 'auditory vection' is a real effect, in which a person perceives self-motion, induced purely through an auditory stimulus (Larsson et al., 2004; Riecke et al., 2009; Väljamäe et al., 2004). Although often difficult to induce, its existence suggests that auditory motion is a distinct, perceived feature. It therefore seems likely that the brain does process auditory motion as a distinct feature of the environment, and not just as a series of discrete locations.

Area PT Although not explicitly mentioned in earlier papers on auditory spatial processing (e.g. Rauschecker & Tian, 2000; Lewis et al., 2000, human fMRI), the Planum Temporale (PT) is likely to be one of the regions involved in auditory spatial or motion processing. Located at the posterior end of the superior temporal sulcus, it has been found to respond more when auditory stimuli are presented from more than one location, even when the auditory stimuli is not attended (Deouell et al., 2007, human fMRI). Differences exist in the definition of PT (Shapleske et al., 1999), and its purpose (with language often being reported as a key stimulus). In Lewis et al. (2000, human fMRI) the belt region around the core and para-belt region (which includes PT) are grouped into a single region 'PAC+'. Figure 4.3 illustrates the resulting simplified pathway. For the purposes of this study, a single auditory region is required to provide the unimodal activity data. With the literature regarding this pathway so mixed, a

simple ROI, either of PT or A1 will suffice.

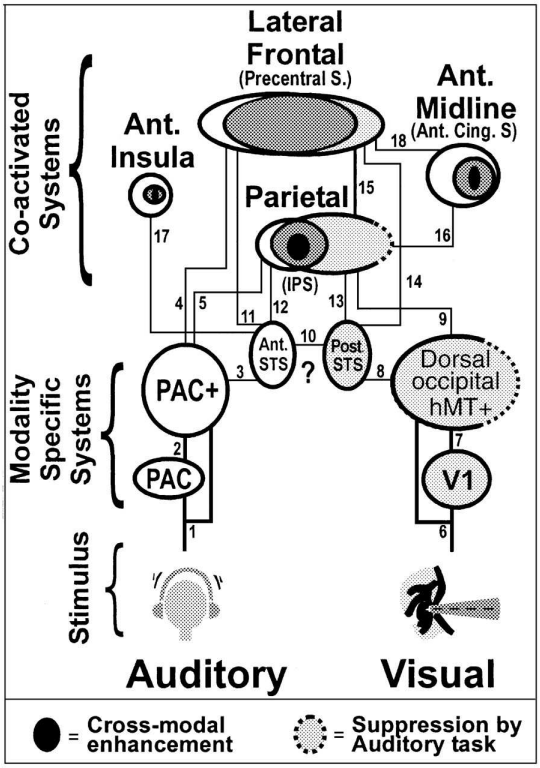


Figure 4.3: Figure from Lewis et al. (2000), illustrating a simplified visual and auditory motion processing hierarchy in humans.

Summary

In summary, the 'where' stream of auditory processing passes from primary and surrounding auditory cortex through a possible hierarchy of regions to the intraparietal sulcus. The 'where' stream could contain auditory motion cues, although no region within the stream necessarily is purely devoted to auditory motion, unlike the visual stream in which regions such as MT are believed to be largely dedicated to motion processing.

Our model generally follows that illustrated in figure 4.3, devised by Lewis et al. (2000). Further complexity certainly exists, but for the purposes of our hypothesis this model will suffice.

4.2.3 Audiovisual motion Integration and Cross Modal Interaction

Types of integration

It appears that three distinct forms of interaction could occur between stimuli of different modalities when processed in the brain (see figure 4.4). The two which are most discussed in the literature refer

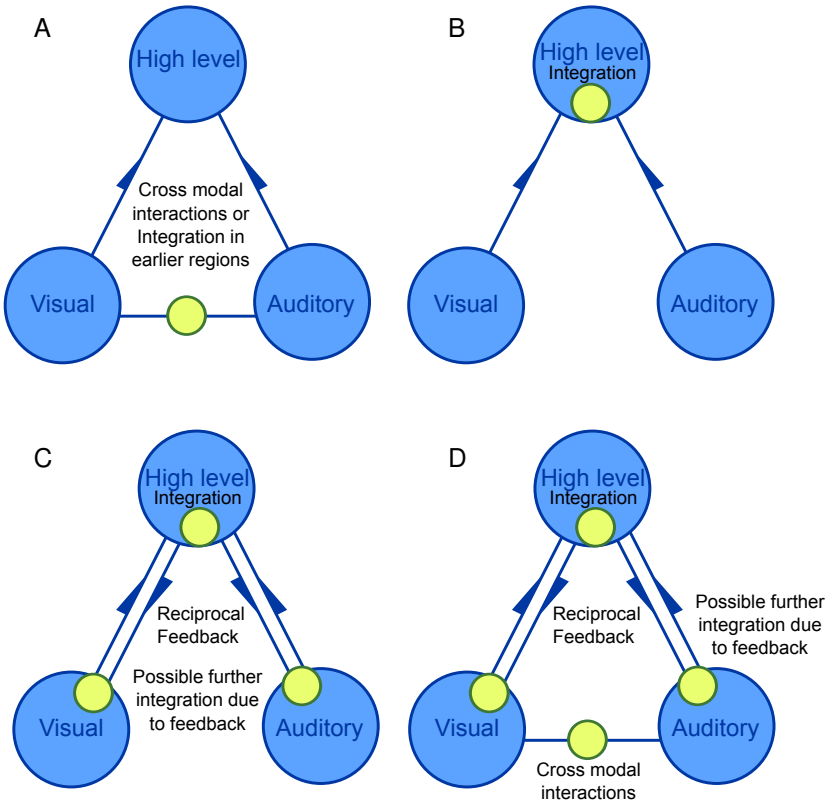


Figure 4.4: Illustration of the different levels at which the two modalities might be integrated. A. Cross modal interactions or integration. B. High level integration. C. Influence of reciprocal feedback on early sensory areas. D. Combination of several mechanisms.

to a choice between high or low-level integration. Alink et al. (2008) and Zvyagintsev et al. (2009) both outline these two hypotheses for the integration of audiovisual stimuli. The 'perceptual hypothesis' suggests integration occurs early in the processing of the two modalities, while the 'decisional hypothesis' predicts the integration will occur after separate processing. In some ways this is similar to the structure described by Lewis et al. (2000), in which information is integrated after being 'derived separately within each modality' (Lewis et al., 2000). Under the 'perceptual hypothesis', within a unimodal sensory area (such as A1), the influence of the other modality will alter the representation of the stimulus. In contrast, high-level cross-modal integration involves the complete combining of features of the two stimuli into a 'supramodal representation' (Lewis et al., 2000). It should be noted that these two mechanisms are not at all mutually exclusive. With evidence for both, it seems likely that both types of cue combination are to be found. Schmiedchen et al. (2012) describe a similar division between crossmodal interactions and multisensory integration, describing cross modal interactions as direct influences of one modality on another without complete integration, while multisensory integration refers to the binding of stimuli from multiple senses. The authors cite the ventriloquist illusion as an example of crossmodal interaction. However, Alais & Burr (2004) make this effect their key exam-

ple of multisensory integration, not of capture. The difference between these studies is the size of the disparity between the modalities; with large disparity and high coherence, the two stimuli may not be integrated. Whether cross-modal interactions exist and precisely what constitutes a cross-modal interaction still needs to be determined (Alink et al., 2011). By using congruent stimuli, we hope that we will be primarily investigating multisensory integration.

The third form of integration is motivated by the observation that many, if not most of the connections within the cortical auditory and visual processing streams are reciprocal (Kaas & Hackett, 2000; Maunsell & van Essen, 1983; Boussaoud et al., 1990; Lewis et al., 2000). This can lead to top-down modulation of activity within low-level unimodal sensory regions (Kamitani & Tong, 2005), leading to apparent cross-modal responses within these regions. Such responses, under this hypothesis, are in fact mediated by high-level integration regions. For example an increase in activity in VIP was detected by Alink et al. (2011) prior to cross-modal dynamic capture (the creation of an illusory auditory motion percept due to cross-modal interaction), although in the paper they argued that audiovisual integration occurs in early motion areas. Deciding which of these models is valid is a challenging task. With current information, it does not seem possible to determine whether the results in Alink et al. (2011) is true cross-modal integration or top-down modulation, or both.

Anatomical and functional connections between unimodal regions One argument against early integration is the relative paucity of connections between these 'unisensory' processing regions. Anatomically there is little evidence of direct connections between the lowest-level visual (e.g. V1, V2, hMT+) and auditory regions (e.g. A1, auditory-belt regions, including PT). Smiley & Falchier (2009, review) did find several connections between auditory and visual areas. They speculate as to a possible connection to A1 from visual areas but the evidence appeared very limited. However, such cortical connections have not been fully explored. Whole brain imaging, such as fMRI, allows functional connections between regions to be assessed. An intrinsic functional MRI connectivity study, looking at the resting state connections between sensory regions found evidence that the primary auditory cortex and hMT+ were functionally linked (Eckert et al., 2008). This would suggest some evidence for such a direct connection between these regions. In contrast, the review by Beauchamp (2005) reports that across several types of stimuli (including motion Lewis et al., 2000) hMT+ has a reduction in BOLD response below baseline in response to auditory stimuli. Further uncertainty is introduced by another study. Alink et al. (2008) found that hMT+ had a greater response when visual and auditory stimuli were congruent, compared to a conflict condition. This clearly suggests that auditory stimuli can have an effect on hMT+. However, attention or some other top-down influence could be causing these changes, a hypothesis supported by the observation that hMT+ had a *lower* response with the audiovisual stimulus (compared to the visual-only stimulus).

The stage at which audiovisual motion is integrated is still uncertain, and it is this author's belief that such integration could occur in more than one region, and by more than one mechanism.

The absence of clear anatomical evidence and the mixed functional evidence suggests no direct connection between these unimodal areas. Therefore the DCM and SEM models used in this study

will not have direct connections between the unimodal regions (A1/PT and hMT+). It should be noted that integration could be mediated by an intermediate integrating region (such as the superior temporal sulcus) which also connects the two regions, in parallel with VIP's integration.

Influence of Stimulus Type

The interpretation of the above experiments strongly depends on the stimuli used. Some stimulus-types may be integrated much earlier than others or may interact differently with attention (Burr & Thompson, 2011).

In this experiment we are looking at the integration of self-motion cues. These are potentially considerably more complex than the planar or single-direction motion investigated in the majority of audiovisual experiments. Most of the auditory studies reported use a single point of sound, which intuitively would be unlikely to emulate self-motion. Auditory vection experiments have found this intuition to be valid, with vection induction more likely when multiple cues are employed (Larsson et al., 2004; Riecke et al., 2009; Väljamäe et al., 2004). The transition as one moves up the auditory 'where' stream, from simple to complex stimuli, reflects the processing of visual motion, with low-level regions responding to any type of motion (e.g. Wall & Smith, 2008) and higher-level regions responding to more specialised motion stimuli (MST responds generally to large-field motion, and VIP to a congruent self-motion stimulus Wall & Smith, 2008). The level at which stimuli are integrated may follow a similar hierarchy, further motivating our choice of complex selfmotion cues instead of the single-object stimuli so far investigated.

4.2.4 The Ventral Intraparietal Area (VIP)

The Intraparietal Sulcus

Before looking at the functional and electrophysiological features of VIP it is necessary to first summarise the nomenclature of the surrounding sulcus. The Ventral Intraparietal area (VIP) of macaque is part of the Intraparietal Sulcus (IPS), which is subdivided into five functional areas: anterior, lateral, ventral, caudal and medial. The intraparietal sulcus in humans is also divided into functionally distinct regions. Silver & Kastner (2009, human, topographic mapping using fMRI) state in their introduction that seven topographically organised parietal areas have been identified in humans, of which six lie within the IPS (described as IPS0 to IPS5, see table 4.1). Choi et al. (2006) report that three areas of human IPS could be identified using cytoarchitectural features (hIP1, hIP2, hIP3, see table 4.2). Clearly we are interested in whether a human equivalent of the Macaque VIP exists in the human IPS. It may be that similar functions are performed in the region, although distributed differently across functional areas. Returning to the Macaque, the VIP is to be found at the fundus of the IPS (Maunsell & van Essen, 1983). Grefkes & Fink (2005) suggest that both VIP and LIP appear to be anatomically relatively similar across primate species, implying that VIP may be found in a similar location in humans. Below we look in more detail at the evidence for this homologue.

Region	Location and Notes	Name
IPS0	Intersection of transverse	
	Occipital sulcus and the IPS	V7
IPS1	Posterior IPS. Same foveal	LIP
	visual field representation. as IPS0	
IPS2	Posterior IPS	LIP
IPS3	Anterior to IPS2, anterior/lateral IPS	(human specific)
IPS4	Anterior/lateral IPS	(human specific)
IPS5	Most anterior, intersection of	VIP
	IPS and Post Central Sulcus	

Table 4.1: Regions identified using fMRI in the IPS, as defined by Silver & Kastner (2009), with the probable macaque equivalent locations.

Choi Region	Macaque Name
hIP1	AIP
hIP2	VIP
hIP3	MIP

Table 4.2: Linking homologous regions in human and macaque Intraparietal Sulcus, as defined by Choi et al. (2006)

Focusing specifically on VIP and its possible homologue in humans, several studies have identified a portion of the human IPS as having a preference (increased BOLD response) for stimuli which are associated with VIP in non-human electrophysiological recording studies.

Schlack et al. (2003) found that the majority of VIP neurons responded to smooth pursuit eye movements (SPEM). This suggests that any equivalent region in humans would also have this association. Konen & Kastner (2008) found that of the regions examined SPEM was most associated with an increase in the BOLD response in IPS5. Neighbouring regions were reported to be more similar than distal regions, a finding reflected in electrophysiological recordings of the IPS in non-human primates and may contribute to the difficulty in assigning homologies between species. The Konen & Kastner (2008) study also found that IPS5 responds selectively to optic flow patterns, the key feature of VIP in macaque (Schaafsma et al., 1997) which we are trying to address. The increase in BOLD response in or near the region IPS5, at the anterior portion of the IPS, has now been widely reported (e.g. Konen & Kastner, 2008; Smith & Muckli, 2010). A more detailed summary can be found in appendix A. Other lines of enquiry also suggest a homologue. The region's response to the tactile stimulation of the subject has been reported in macaque (Colby et al., 1993), with the neurons responding to both this and visual stimuli. In humans a study by Sereno & Huang (2006) found the anterior region of the IPS they had

identified as possibly being a homologue of the region had an increased BOLD response to air-puffs to the subject's face. This set of results does seem to suggest that VIP has been partially preserved between primates, and allows predictions regarding the integration of cues (including audiovisual integration) and the relevance of self-motion cues in the region in humans, based on the results of recordings in non-human primates.

It finally should be noted that the homologue of human VIP may actually consist of at least two subfields (Lewis et al., 2000, human fMRI). Such a distinction is very difficult to make in an fMRI study, and the details of its division are still preliminary, so the region is regarded as one functional area for the purposes of this study.

Anatomical Connections to VIP

The anatomical evidence for connectivity is largely based on tracer-studies in non-human primate. Maunsell & van Essen (1983) found that MT was connected to VIP through reciprocal connections, a finding confirmed by Boussaoud et al. (1990). Anatomical evidence of VIP's connection with A1 was less clear. Lewis et al. (2000)'s review summarises the connections to VIP as being from three areas associated with auditory processing: the caudal temporal opercular, the temporoparietal area and the polysensory temporal parietal occipital complex. The caudal temporal opercular probably matches, or is close to the location of A1's core or belt region, and so suggests there may be direct connections. They also report that almost all connections they have observed are reciprocal. Although other auditory (and visual) areas directly innervate neurons in VIP, it seemed reasonable for A1 and hMT+ to be candidate regions for this study.

VIP Activity: Integration

A large number of studies have focused on VIP, leading to a considerable number of hypotheses regarding its purpose and function. Electrophysiological studies have found that cells in VIP respond to motion in four modalities: visual (e.g. Zeki et al., 1991; Colby et al., 1993), vestibular (Bremmer et al., 2002b), somatosensory (Duhamel et al., 1998) and auditory (Schlack et al., 2005). The neurons' reference frames lie on a continuum between the two 'input' reference frames (head-oriented or eye-oriented) (Schlack et al., 2005, in Macaque). Intriguingly this spread of reference frames is true for neurons responding to either modality, so neurons which respond to auditory stimuli can be in the eye-centred coordinate system and vis-versa. This complexity is sidestepped in this study by requiring participants to fixate on a stationary fixation cross.

Much electrophysiological research has focused on the integration of congruent visual and somatosensory stimuli in VIP neurons (Duhamel et al., 1998) and visual and vestibular motion cues (Bremmer et al., 2002b). However we are interested in VIP's integration of visual and auditory stimuli, and its preference for self-motion congruent optic-flow. Of particular relevance is a study by Schlack et al. (2005), in which visual and auditory stimuli were presented in different spatial locations to a macaque, while recording from neurons in VIP. Schlack et al. (2005) reported that the great majority of cells re-

sponded to both the visual and auditory stimuli, the first time auditory stimuli have been reported to elicit a response in VIP. Most interesting was the congruence in the auditory and visual receptive fields. For most of the bimodal neurons, the receptive fields of the two stimuli heavily overlapped.

Activity: Visual Self-motion

Early studies in VIP found a preference for visual motion (Colby et al., 1993). Work by Frank Bremmer (Bremmer et al., 2002a) and others (Wall & Smith, 2008) followed the hypothesis that, because VIP received strong inputs from MST, an area known to respond to self-motion stimuli, VIP would also have a preference for *self-motion* congruent cues (see Bremmer (2005) for review). VIP neurons have also been found to be receptive to the location of the centre-of-expansion (Bremmer et al., 2002a; Zhang & Britten, 2004) A key observation which motivated the choice of a heading discrimination stimulus and task. Interestingly, most neurons appeared to respond most to centres of expansion near the direction the head was facing, similar to the responses of MST neurons (Duffy & Wurtz, 1995). These observations suggested the use of an almost-forward self-motion cue would be the most successful stimulus.

Integration of visual and auditory motion stimuli

The obvious implication of these two features is that VIP contains neurons receptive to auditory self-motion stimuli and integrates these auditory cues with visual self motion cues. This inference has not yet been proven, but indirect evidence suggests that VIP may indeed integrate both auditory and visual self-motion cues. The first evidence was from a study of people suffering lesions to the right parietal cortex. It found that sound-motion perception was compromised in those cases (Griffiths et al., 1997). It is likely other regions were also compromised, and lesion studies are often difficult to interpret. The increasing use of fMRI has allowed this topic to be investigated with more precision. The first study to investigate was Lewis et al. (2000) in which a sense of auditory motion was induced by a change in the IID of a square wave. This was paired with a moving visual stimulus. They found that activity in the intraparietal sulcus (IPS) was increased for both stimuli and cross-modal enhancement occurred in the region.

Since that study several other fMRI studies have been conducted to look specifically at auditory motion responses which might indicate if VIP is involved in processing or integration of such stimuli. Bremmer et al. (2001) used conjunction analysis to show that areas within the intraparietal sulcus responded to both auditory and visual motion stimuli. Some studies purport to show incredibly large regions to be responding to auditory motion. For example Poirier (2005) report huge swathes of parietal cortex respond to auditory motion. No functional imaging study has yet looked at the integration of self-motion congruent visual and auditory stimuli.

Other regions

It is important to note that area VIP probably does not act alone in integrating visual and auditory motion stimuli. As mentioned above, Poirier (2005) found large numbers of regions responding to an auditory

motion stimulus, while Lewis et al. (2000), using a cross-modal enhancement paradigm, report that, besides IPS regions, integration may occur in the anterior insular and anterior cingulate sulcus. Baumann & Greenlee (2007) report that a moving vs static auditory stimulus induces an increase in activity in several regions. As expected, the main area reported is the superior temporal gyrus. Of more interest are the areas detected when a combine audiovisual stimulus was used. They report that four areas are involved in audiovisual motion processing: the superior parietal lobule (SPL), the supramarginal gyrus (SMG, area 43), the intraparietal sulcus (IPS) and the superior temporal gyrus (STG). This follows much earlier electrophysiological research reporting visual and auditory motion responses in the temporal lobe (e.g. Benevento et al., 1977), a very early study which also investigates their interaction). Many other studies have found these and several other regions to be responsive to both auditory and visual motion (see Britten, 2008, for review).

Summary

The electrophysiological and human imaging studies performed to date show VIP responds to auditory motion (although the concerns expressed by Smith et al. (2007) regarding the choice of control stimuli should not be ignored), and many studies also show it responds to visual self-motion stimuli. Absent from the literature are studies looking at the integration of both auditory and visual self-motion stimuli in the region.

Cross-modal interaction may occur at early stages of processing, or may be driven by top-down interactions. However, even if some features of visual and auditory motion are partially integrated prior to reaching VIP, VIP seems like a reasonable candidate for the full integration of the self motion components of the stimuli. This research sets out to test if VIP is a region which integrates the visual and auditory self-motion cues presented.

4.2.5 Coherence and Connectivity

Summary

Previous research (see introduction to Chapter 3) has found that humans and other animals integrate cues from a wide variety of modalities in statistically optimal ways (e.g. Bayesian integration). However, the underlying neural processes which perform this integration are not well understood. Several competing theories and hypotheses exist and have been investigated with both electrophysiological recordings and whole brain imaging. Cell-recording studies are able to test whether the same cell responds to stimuli from different modalities, and how the population of neurons responds to changes in coherence, while imaging studies attempt to investigate changes in connectivity by looking at how the the correlation between regions changes with changes in coherence.

Linear summation model

Beauchamp et al. (2010) used Structural Equation Modelling to analyse connectivity between a pair of unimodal 'early' areas (Visual and Somatosensory) and the Intraparietal Sulcus (IPS), believed to be an area involved in integrating visual and somatosensory stimuli. They refer to two competing theories of integration, 'linear summation' and 'weighted connections'. Linear summation is a theory proposed by Ma et al. (2006) to explain how neural circuits can perform Bayesian integration, through simply adding neurons' firing rates. Ma et al. (2006) predict that the poisson nature of neural activity, combined with the population coding of stimuli, will produce activity in the integrating population similar to a Bayesian 'optimal' result.

Regarding correlation analysis (such as structural equation modelling), the linear model predicts that (given they have the same level of activity) the two input populations are represented equally in the integration population. Therefore we should not find correlations changing depending on stimuli coherence (besides changes in overall activity) (Beauchamp et al., 2010).

Weighted connections model

The divisive normalisation model (Ohshiro et al., 2011), a more recent model (described in section 3.3.5) also attempts to explain population integration, but also predicts changes in the combination-rule in response to cue reliability, reported by Morgan et al. (2008). This model could produce the same results as the weighted connections hypothesis, the second model proposed by Beauchamp et al. (2010).

In both models, an increase in coherence or reliability of a stimulus is predicted to cause an increase in the correlation between the relevant unimodal brain area and the integration area. Beauchamp et al. (2010) found that correlation between regions increases with the stimulus coherence, suggesting that the 'weighted connections' model is more likely than the linear-summation model. It is this connectivity analysis we attempt to reproduce in our self-motion paradigm. Studies in other fields have also found coherence modulates connectivity, for example, Kreifelts et al. (2007), a similarly motivated study, found changes in functional connectivity between regions associated with speech comprehension when coherence was varied. Nath & Beauchamp (2011) investigated the perception of speech, by using auditory and visual cues of words being spoken. Unlike Beauchamp's earlier work (Beauchamp et al., 2010), both the auditory and visual cues were degrading by reducing stimulus coherence and not by adding new stimuli (such as noise). However, they did not report unimodal behavioural task performance. If the stimulus had insufficient coherence (i.e. was impossible to perceive), then it is difficult to determine whether the effects reported are due to integration or the failure of the low-coherence processing area to encode any meaningful component of the stimulus. They estimated parameters separately for each participant, and found the correlation between the unimodal region with the greater coherence was more connected to the superior temporal sulcus (the putative integration region).

Superadditivity as an indicator of integration

fMRI studies (e.g. Lewis et al., 2000) and many electrophysiological studies (e.g. Meredith & Stein, 1986) have found multi-modal enhancement (superadditivity). This is where a cell or region has a greater response to the combined, bimodal input than the sum of its responses to the two unimodal stimuli, presented separately. It is used as an indicator of multimodal integration.

Of the two models mentioned earlier, the linear-summation model does not predict any super-additivity, while the normalisation model does. Ma & Pouget (2008) (presenting the linear-summation model) argue that super-additivity at a neuronal level will cause saturation, degrading the accuracy of integration, and that reports of super-additivity in both electrophysiological studies and fMRI studies are due to either a minority of cells or the misinterpretation of fMRI data. A study since then, by Gu et al. (2008) reported sub-additive weighting in the summation of visual and vestibular cues in MST.

The reason for this difference may be the coherence of the stimuli. The review by (Stanford & Stein, 2007) suggests that the emphasis on super-additivity in the electrophysiological literature is due to the use of low-coherence stimuli. This is matched by the normalisation model, in which superadditivity mainly occurs at low coherence (when it is most needed), and becomes sub-additive at high coherence, when it is least required, and as observed by Ma & Pouget (2008), could cause saturation and signal degradation.

In summary, this author believes superadditive multimodal enhancement is a real effect and should be considered an indicator of integration. Because audiovisual superadditivity has already been the focus of research in VIP (Lewis et al., 2000). The following experiment was not designed to test for superadditivity. This allowed us to focus the experimental design on investigating the effects of coherence.

4.2.6 Non-stimulus responses

Attention

One of the most clear examples of top-down modulation is the effect of attention on cortical sensory regions. For example, Kamitani & Tong (2005) found attention could modulate activity in early visual areas such that ensemble activity would encode the attended features in the stimulus. In our experiment, the participants in the low-coherence group reported not distinguishing between the two coherence conditions, so were not consciously attending to one stimulus or the other, but attention changes may still be occurring. It is possible therefore that the BOLD responses reported in the experiment are modulated to an extent by attention, for example, if the visual stimulus is offering greater accuracy (through an increase in coherence), one would expect attention (conscious or not) to be directed towards this modality. We believe this may be an important component of cross-modal integration, and so we did not want to disrupt this component through the use of a distractor task.

The lack of control of attention is a reasonable criticism of the study. Alternatively we could have made the direction dimension of the stimulus irrelevant from the participant's point of view, and instead,

asked the participants to attend to a different aspect of the stimulus (for example its speed). This would have allowed us to state that the effect was independent of the attentional focus. This design would mean the participant's decisions couldn't be used to decoded the fMRI data in a comparison with the true stimulus classification, an area I was hoping this experiment would have allowed me to investigate.

The alternative design would be to include trials with and without attention in those modalities of importance. This would allow the effect of attention to be investigated by comparing between those conditions. This design has the drawback of requiring many more trials, or fewer samples for each condition. It was felt the experiment had already been weakened by compromising between several research questions, so this additional comparison would not be wise to include.

Prediction error or Surprise

den Ouden et al. (2010) suggest that one important predictor of the BOLD response in sensory areas is to do with participants' expectations about the stimulus. An unexpected stimulus, they and others report, leads to an increase in the BOLD response in the relevant sensory region. They also find that the connectivity between a visual area and the dorsal premotor cortex (an area associated with making decisions regarding actions) is modulated by this prediction error. Our experiment doesn't allow this feature to be investigated, as we cannot estimate, from the experimental design, the expectations of the participants.

This effect could be responsible for changes in connectivity, for instance by an increase in connectivity to the region associated with low coherence due to the increased number of prediction errors the region might make.

4.2.7 Summary of Methods

Several methods exist for investigating both connectivity and integration using fMRI; adaptation analysis, cross-modal MVPA and connectivity analysis.

Adaptation

fMRI adaptation is a common method for studying such questions (Krekelberg et al., 2006) (section 4.3.12). It allows sub-populations of neurons within single voxels to be indirectly isolated. In our case we will attempt to identify the sub-population which responds to motion in one modality (e.g. visual motion), and see if this population also responds to the same direction of motion in a different modality (e.g. auditory motion). One caveat to this hypothesis is the finding that some neurons in the region prefer visual and vestibular stimuli of opposing directions (Bremmer et al., 2002b). The detection of cross-modal adaptation may be difficult if the audiovisual neural responses have a similar conflict to direction as the visual/vestibular responses. The response to the opposing direction may become equally adapted (if an equal number of neurons prefer opposing directions of motion). The above assumptions depend on adaptation occurring in putative 'human VIP' in response to the visual stimulus. A figure in a paper

by Bremmer (2005) appears to indicate such adaptation does occur (figure 4.5, as the response of the illustrated cell initially peaks at over 60 Hz for the first 500ms, then falls to below 40Hz for the second half of the stimulus. This author however, knows of no study using adaptation and self-motion stimuli, within this IPS region.

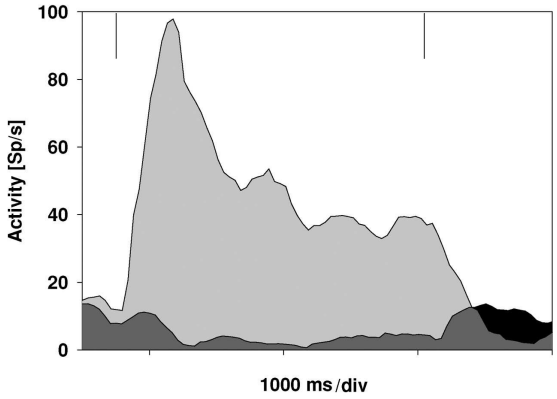


Figure 4.5: Figure from Bremmer (2005). Reproduced here to illustrate the adaptation effect in a VIP neuron. Original caption: Neuronal responses for expansion and contraction stimuli: The spike density curves show the data for testing a cell with stimuli simulating forward (expansion; light grey) and backward (contraction; dark grey and black) motion. The vertical lines indicate stimulus on- and offset. The cell clearly preferred simulated forward motion.

Cross Modal MVPA

An alternative method, which will indirectly indicate whether integration is taking place is to look for cross-modal multivoxel decoding. Multivoxel decoding or pattern analysis (MVPA) is a method which uses the inhomogeneity in response to a stimulus across a region to provide extra predictive information about the stimulus. If decoding of motion direction is possible in putative human VIP from either stimulus, one could test whether a decoder trained on one stimulus modality can predict the direction using the responses to the other modality. If cross modal decoding is possible, then those neurons which respond to the same direction are likely to be grouped together - and may be the same neurons. Such clustering could also indicate a functional relevance, although would not be as direct as adaptation analysis. A third method involves the invoking of superadditive responses (Lewis et al., 2000), as indicators of multimodal integration. The experiment was not designed to exploit this effect however.

Connectivity

Using a similar method to Beauchamp et al. (2010) we investigated the effect of stimulus coherence on the connectivity between regions. The key hypothesis was that the sensory region receiving the stimulus with the greatest coherence will have the greatest correlation with the integration region because the population input from that region will be ‘weighted’ more strongly than from the less coherent region.

This hypothesis follows previous fMRI connectivity research and models such as the normalisation model described in the last chapter. Other research and models suggest that this effect does not exist, or is not necessary for the Bayesian integration of stimuli.

Neural activity and behaviour

Beyond investigating integration and connectivity, the study may also answer some further questions. First, whether decoding stimulus direction (in either modality) is possible; failure to decode direction would suggest this stimulus needs refinement for use in fMRI studies. Second, which is more successful: the decoding of the stimulus or the decision made? If the behavioural decision made is a better predictor of the activity in the region than the stimulus, then the activity would be a window on perception. If the converse occurs and the activity is more related to the stimulus, one must ask why the perception of the stimulus is different from the neural activity. Thirdly, how does coherence influence where decoding is most successful? One would expect a greater coherence to allow more successful decoding, but the reverse may occur through the effect of attention or superadditivity. Finally, the study is expected to duplicate previous results, such as the overlap of BOLD responses in putative human VIP, and other indirect indicators of an integration area.

Note that there is considerable uncertainty about whether humans actual have a single equivalent region to VIP. Section 4.2.4 summarises the connection between the currently identified fields in human IPS and what may be the equivalent areas in non-human primates, including region VIP.

4.2.8 Summary

This experiment will expand on the psychophysics in the previous chapter to investigate in more detail how the unimodal audiovisual self-motion cues are integrated. In particular I hypothesise that the putative human Ventral Intraparietal Cortex responds to both stimuli, and combines them at a neural level. This leads to the hypothesis that cross-modal adaptation will be found in VIP, and cross-modal classification. Second superadditivity is hypothesised to be found in VIP, i.e. the BOLD signal recorded in the bimodal condition will be more than the sum of the responses in the unimodal conditions. Our other hypotheses concern the way in which the reliability or coherence of the two stimuli changes the way in which the cues are combined. Following from previous models of integration, we hypothesise that the most coherent stimulus will lead to the highest connection between the relevant unimodal region and the bimodal region. This connection weight hypothesis will be tested by using both dynamic causal modelling and structural equation modelling; two methods for investigating the correlation and functional connectivity between regions. This collection of tests and hypotheses regarding the integration of self-motion will hopefully shed some light on this complex field.

4.3 Methods

4.3.1 Participants

Fifteen participants took part in the study (one took part twice, once in the low-coherence group and once in the high-coherence group), aged 19-35 (mean 24.0, standard deviation 4.4, 7 female). All reported normal or corrected to normal vision, and none reported hearing loss. Hearing or sight loss significant enough to affect the experiment would be detected during the training prior to the scan. Participants were recruited through the University of Glasgow's Centre for Cognitive Neuroimaging (CCNi) subject pool database. The study was approved by the ethics committee of the CCNi.

The two groups are described below. Because the study required long periods in the scanner, a difficult psychophysical task and several tools (visual and auditory stimuli, eye tracking and the button box), four participants had to be rejected completely, and two partly, due to participant fatigue or equipment failure. These issues are described below.

Low Coherence Experiment

In the first part of the experiment (with low coherence stimuli) ten participants took part. A failed amplifier meant sound was not delivered reliably to one of these participants. Data from another participant was rejected as they fell asleep and a third participant was unable to satisfactorily complete the visual component of the task (failing to reach a threshold of less than 100%). This appeared to be a problem of understanding the task, and not due to poor visual acuity. This left seven participants, aged 20-34 (mean 24.1, standard deviation 4.4, 2 female). The final two had their bimodal sessions disrupted by the reversal of the left-right channels of the amplifier causing them to experience an unwanted conflict condition in the bimodal sessions.

High Coherence Experiment

In the second part of the experiment (with high coherence stimuli) six participants were scanned (one of whom had already taken part in the low-coherence experiment). One had very excessive head motion and was rejected. This left five subjects for analysis, aged 19-35 (mean 24.8, standard deviation 6.2, 3 female).

Sample Size

Clearly five participants is insufficient for a study of this sort, but it may still be possible for the two populations to be combined for analyses such as the dynamic causal modelling and structural equation modelling. Strong classification results across five participants would also be considered sufficient, at least to give an indication of a wide-spread effect. Any conclusion reached with a population of this size will need additional, larger, studies to confirm any results.

Ideally additional scans should have been taken, but resource limitations meant that the scanner was not available. Indeed, only a sample of twelve participants was originally planned, but the technical problems described above meant that we increased the number of participants. Future studies should try to negotiate greater flexibility from the research centre, and preferably the costs of scans should not be so critically constrained.

4.3.2 Apparatus

fMRI scanning

Blood-oxygen level dependent (BOLD) signals were recorded using the University of Glasgow, CCNi fMRI scanner (Siemens 3T TIM Trio, using a 32-channel head coil). We acquired 371 volumes in each unimodal session, and 622 in the localiser session. Each volume had 34 slices, with slices of 2mm (gap 0.5mm), with $2mm \times 2mm$ voxels. Almost the whole cerebral cortex was acquired, with only the most dorsal portion of the precentral gyrus and the most ventral part of the inferior temporal gyrus outside the imaged volume.

The anatomical scan was a standard T1-weighted MRI sequence (3T. TR=1900ms, TE=2.52, FA=9°, $1mm^3$ isotropic voxels) taken during the same scanning session.

Eye tracker

MST is known to respond to eye-movements (Newsome et al., 1988) and VIP neurons appear to have a mix of head- and eye-orientated coordinate systems, so precise fixation is important to avoid confounds caused by eye movement.

To confirm the participants successfully fixated during the study, and to ensure that there was no difference in eye-movement between stimuli, the location the participants were fixating was monitored using an eye tracking camera and software (Viewpoint Eye-Tracker, Arrington Research, Inc., Scottsdale, AZ). The eye tracker was recalibrated before each session during the experiment. A parallel port interface between the stimulus computer and the eye-tracker was installed to allow the time stamps in the eye-tracker log files to be synchronised with those for the visual stimulus log. The variance in the location of fixation between conditions was compared using a custom MATLAB script.

Headphones and amplifier

MRI-compatible headphones (NordicNeuroLab, Bergen, Norway) were used to present the auditory stimuli, and attenuate the scanner sound. Additional padding was firmly placed beside each ear to further muffle scan noise and minimise head motion. Prior to scanning, a standard 1kHz sine wave sound was used to test the earphones and adjust them to have equal 80dB output. The perceived volume may vary depending on each participant's hearing. However, both participant adaptation and the calibration during the anatomical scan would have normalised such differences.

Projector and screen

Visual stimuli were projected with an LCD projector onto a screen placed behind the subject. This was viewed using a mirror placed in front of the subject, angled at 45°. The visual stimulus was produced using Vizard 3.0 (Worldviz, www.worldviz.com). The screen appeared 110cm from the participant's eyes, and extended 18.1° horizontally and 14.0° vertically.

4.3.3 Stimuli

Visual

The stimulus was similar to that in the psychophysics experiment, with disks positioned in 3d in such a way that they formed a plane. The whole background, behind the disks, was light grey. Each disk translated across the plane with a fixed speed and trajectory and had a limited lifetime of between zero and one second (uniformly distributed). To ensure that optic flow was not disrupted by the appearance and disappearance of the disks, they were added and removed by gradually increasing and reducing their opacity over a period of 0.3s (thus some did not reach full opacity). The view point was above the plane, at a virtual height of 1.8m.

At 100% coherence the dots all translated in the same direction (across the plane), giving the appearance that the subject was moving forward over the plane of dots, however the motion was not from directly ahead, but was rather angled 10° to the left or right of a line facing ahead. Note that this equates, with a field of view of 88°, and a screen of 18.1° across, to a stimulus in which the centre of expansion was 2.46° to the left or right of midline (marked by a fixation cross in the centre of the screen).

The main differences were that it consisted of more disks (approximately 110 ± 15 on screen at any time) which were each half as wide as those in the psychophysics experiment (0.1m across, 38 arc minutes maximum diameter from participants perspective), and didn't use anaglyph rendering. Also participants were strongly instructed to fixate on the cross during the experiment to ensure retinotopic regions remained stable and eye-motion did not cause signal noise or influence motion processing in VIP. Figure 4.7A illustrates the stimulus, with the central fixation cross.

Auditory

The sounds used were almost the same as in the psychophysics experiment. It became apparent upon analysing the psychophysics behavioural data from that study, that some of the sounds were easier than others, for the same 'coherence' level. This was unavoidable given that the location and volume of the sounds changed and the precise segment of the recording used varied between tracks. One way to avoid this would have been to use the same or very similar recordings each time, but this would have been very vulnerable to over-learning in such an extended experiment. Instead, the error-rate for the twenty-five tracks was plotted (see figure 4.6), and the bottom (easiest) six and top (most difficult) four were removed, leaving a set of fifteen tracks. These still had some variation ($22\% \pm 5\%$ accuracy on average), but were similar enough to allow the experiment to be unaffected by this variance in the stimulus.

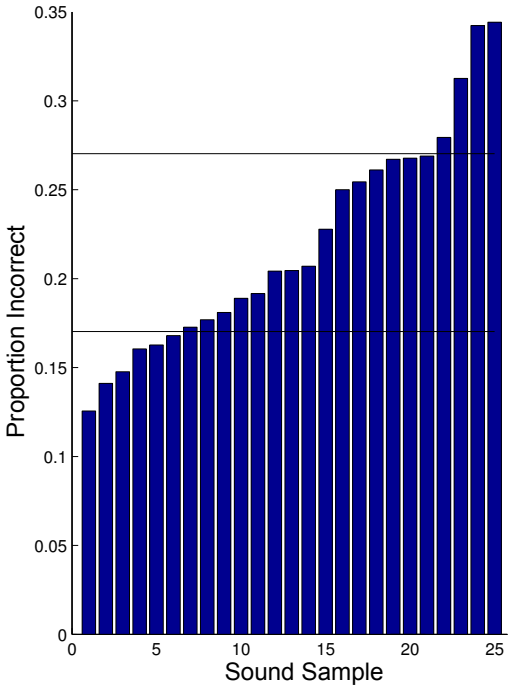


Figure 4.6: Selecting the sounds with a medium level of difficulty.

The other change in the auditory stimulus was to make the sounds considerably shorter. Each audio track was reduced to 3.75s to allow as many trials as possible to be included in each session.

4.3.4 Procedure

Instructions and Training

As in the psychophysics experiment, a series of instructional and training examples were given (prior to entry into the scanner) to ensure the participants understood the task, and to reduced the effect of learning that occurs during the initial performance of the task. The thresholds, instructions and baseline sessions were the same as in the psychophysics experiment (described in section 3.2.4). These were necessary to determine the thresholds for each person for the one-up-two-down (71%) and one-up-four-down (84%) thresholds.

Once the participant was in the scanner, two further preliminary sessions were performed. To maximise the time spent imaging in the scanner, these were conducted during the anatomical scan. This also meant that the scanner would be producing similar sounds (although not the same, due to the different pulse sequences used). These two sessions allowed the one-up-two-down and one-up-four-down thresholds to be re-estimated in the new environment. In each session, 12 trials of each of the four conditions (auditory/visual at high/low coherence) were conducted. New thresholds were calculated on-the-fly from these trials and used in the main experiment. The values were recorded by the experimenter, to

confirm the stimuli were being correctly perceived.

Trial types

Unlike the psychophysics experiment the trials were shorter (3.75s long) and the way in which the participant was required to respond was altered to avoid the manual response choice from confounding the fMRI data. Unlike in the psychophysics study, participants were not permitted to respond before the end of the stimulus. Once the stimulus was completed an arrow would appear in the centre of the screen (figure 4.7B) pointing either leftwards or rightwards. Participants were required to press the button on the button box if the arrow faced in the same direction as they had been moving in. This solution meant that the manual responses were the same for both directions, avoiding confounds.

Occasionally no stimulus would be shown (a rest trial). This would consist of just the fixation cross (figure 4.7C). This would last the same length of time as a normal trial and would be followed 50% of the time by the continued presentation of the fixation cross or half the time the cross would be replaced with a disk (figure 4.7D). The participants were required to press the button upon seeing this. This was to ensure the rest period had the same number of button presses as the stimulus trials, and to ensure attention remained roughly similar across trials, although clearly the attended stimulus would be different.

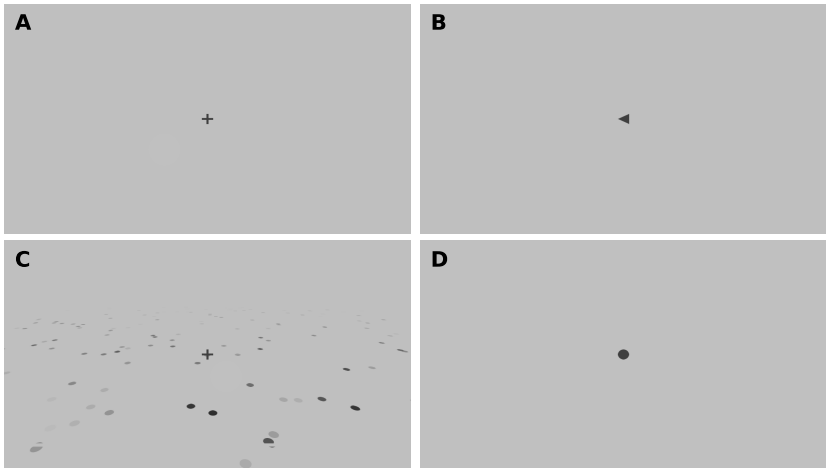


Figure 4.7: Examples of the visual stimuli. A. During auditory or rest trials just the fixation cross is visible, B. After the auditory or visual stimuli, the software displays an arrow, waiting for a response (i.e. press the button if the arrow is facing in the direction you were travelling in), C. Visual stimulus, the dots appear to form a plain across which the participant is moving. D. In half of the rest trials, the cross becomes a dot, indicating that the participant should press the button.

Unimodal sessions

The unimodal sessions were devised to test several hypotheses. First that the direction of motion could be decoded for either type of stimulus. Second, that the perceived direction of motion would be a more accurate regressor of stimulus motion than the actual motion. Third we believed that adaptation would occur for stimuli of the same direction (of the same modality), and that we would find adaptation across modalities in the integration region (VIP). Fourth, if decoding within modalities succeeded, we hypothesised that the same decoder would be able to decode the stimulus direction in other other modality.

In the low coherence group three unimodal sessions were completed by each participant. This meant the experiment lasted a considerable length of time. Combined with eye-tracker calibration, the challenging nature of the task and the relative complexity of the stimuli the experiment length was reduced by using just two unimodal sessions for each participant in the high-coherence group.

The low coherence group were given the 84% correct (4-up-1-down) stimuli, while the high-coherence group were given 100% (completely) coherent stimuli. This change was made due to concern that decoding of the stimuli was failing.

Adaptation Because the stimulus was an event related design, in which adaptation was being investigated, the order of the stimuli was vitally important. There were five types of stimuli: Null(0), Leftward Visual(1), Leftward Auditory(2), Rightward Visual(3), Rightward Auditory(4).

Given these five stimulus types, there are twenty-five combinations of current and preceding trial. Each combination needs to be tested to see what effect each stimulus has on every other stimulus' response. However, the preceding trial's BOLD response may be affected by the trial two-back. This needs to be taken into account and should be counterbalanced. When this two-back counterbalancing is included we can see there are 125 combinations of trials ($000, 001, 002,\ 003, 004, 010,\ 011$.....$434,\ 440, 441, 442,\ 443, 444$). In general, if the experiment looks back $n-1$ trials, taken from an alphabet of size k there will be k^n *tuples*. To ensure the experiment is efficient, a sequence of 127 trials should be devised which includes all these triplets within it. For example, the sequence may begin 0102034... and therefore includes $010, 102, 020, 034$. Aguirre et al. (2011) note that Nicolaas Govert de Bruijn and Tania van Ardenne-Ehrenfest discovered and investigated such sequences and found that the number of possible orderings are:

$$N_{seq} = \frac{k!^{k^{(n-1)}}}{k^n} \tag{4.1}$$

Where k is the number of letters in the alphabet and n is the number of trials in each tuple. In our case, $n = 3$ and $k = 5$, which means $N_{seq} \approx 7.63 \times 10^{49}$. In theory, any of these orderings would be suitable. However, the nature of fMRI signals means some are more likely to detect the features of interest.

de Bruijn cycles fMRI BOLD signals are temporally-filtered approximations of aspects of neural activity (see section 1.1.2 for more details). The filtering occurs at high-frequencies (above 0.1Hz) due

to the sluggish haemodynamic response. At low frequencies (below around 0.01Hz) auto-correlated noise from both physiological and non-physiological sources (Zarahn et al., 1997) must be avoided, and this is usually achieved by applying a high-pass filter to the data (standard in SPM's GLM analysis). This leaves a band of frequencies between 0.1Hz and 0.01Hz in which our signals of interest must lie.

In a simple two condition experiment, one can easily select the block length to fit within this band of frequencies. To achieve the same for the adaptation experiment is more complicated. Some transitions are expected to produce greater BOLD responses than others. To maximise our chance of detecting the difference between these trials we should place them so that the predicted BOLD signal follows a sine wave (or waves) within the 0.1 Hz-0.01 Hz frequency range. Failure to do this could risk the responses of interest being attenuated. For example, if every other trial was expected to have a strong adaptation result, and the trials are only 4.5s long, we would have an expected BOLD signal of 0.11Hz, which would be somewhat weakened by the haemodynamic lag. Similarly, if we have all the trials we expect to have a high BOLD response for the first 3 minutes, and all the low-response trials in the second 3 minutes, our signal would have a frequency of only 0.003Hz and would be heavily attenuated by our high-pass filter.

As mentioned above, there are 7.63×10^{49} possible orderings of trials. Some of these will have more of their expected signal within the desired range of frequencies than others. To select those with the most appropriate frequency responses we made use of the recent work by Aguirre et al. (2011), where they guided the construction of de Bruijn cycles by grouping the transitions into different expected BOLD signals, and then selecting the group (and therefore transition) with the nearest matching BOLD signal to allow the expected BOLD response to follow a sine wave of the desired frequency.

We used the software tool Aguirre et al. created (debruijn[1]) to calculated these optimised sequences to generated the order of trials for the unimodal conditions. Figure 4.8 illustrates how the expected BOLD response will have frequencies within the range the fMRI scanner is sensitive to.

As will be mentioned later, the debruijn cycles used were accidentally incomplete, with the last n-1 trials missing, meaning that only 125 trials were included in each session, not 127. This means the counterbalancing was not quite complete, but is unlikely to pose a serious problem to the experimental results.

Bimodal sessions

The bimodal sessions were developed to test for changes in connectivity between the unimodal and bimodal regions in response to changes in coherence.

The bimodal sessions consisted of 126 trials. These were grouped into fourteen blocks to maximise the signal strength. Each block was of one of two types: Either the visual stimulus had a greater coherence than the auditory one, or vis-versa. In the low-coherence group, the coherences were set at either the 84% accuracy or 71% accuracy thresholds, while in the high-coherence group the coherences were set at 100% and 84% accuracy thresholds. The stimuli were either leftward or rightward, with

[1] Used v1.5, cfn.upenn.edu/aguirre/wiki/public:de_bruijn_software

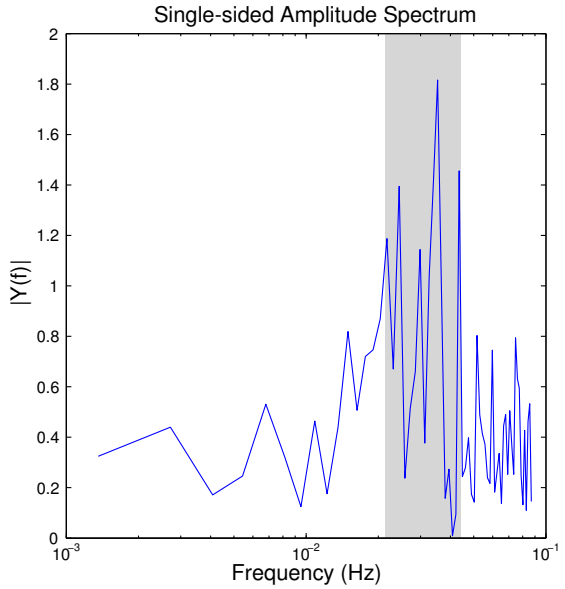


Figure 4.8: Predicted frequency spectrum of the BOLD response, using the optimised de bruijn sequence. For this example, the shaded area indicates the range of frequencies the algorithm was tasked with enhancing.

three of each in each block. Each block had six trials, followed by three null trials. There were seven of each type of block in each session (see figure 4.9). The trial length, response and stimuli were the same as described above for the unimodal sessions.

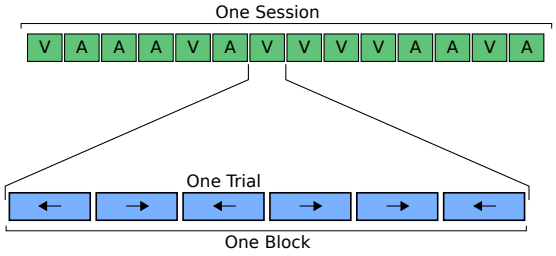


Figure 4.9: Bimodal session design. The session is split into 14 blocks, 7 in which the visual stimulus is of greater coherence than the auditory and 7 blocks of the reverse. Within each block there are six trials, three in each direction.

4.3.5 Functional Localisers

We needed to functionally localise hMT+ (and its subfields MT and MST) and VIP. Other regions of interest (Plantum Temporale and V1) we assumed we could anatomically segment, using the T1-weighted structural MRI session data. Appendix A describes some relevant localisers used previously

and the motivation for the localisers used in this experiment.

Design

The functional localiser consisted of 60 blocks in a pseudorandom order to ensure there were ten blocks of each of the six conditions. These six conditions (see Figure 4.10) were; 1. No dots (just a blank screen with the fixation cross), 2. Spiralling dots (consistent selfmotion). 3. nine separate spirals (inconsistent with selfmotion). 4. Static dots (no motion). 5. Spiralling dots (left hemifield). 6. Spiralling dots (right hemifield). Each block lasted 15 seconds with just a 0.425s blank between blocks. Every 5th block was followed by a question for the distractor task (lasting 5 seconds).

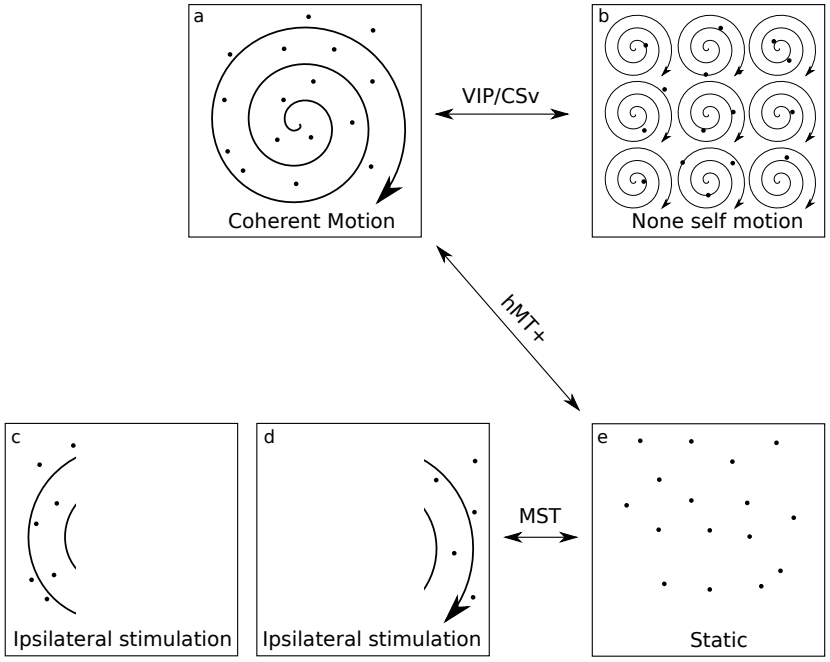


Figure 4.10: Localisers used in this study. A, Spiralling dots; B, Nine separate spirals; C and D, one hemifield spiralling dots; E, Static dots.

Stimuli

900 dots were on screen for all the full-screen stimuli. In all conditions, each dot had a diameter of 6 arc minutes. In all motion conditions the dots moved at a fixed $6.77^{\circ}/s$. This meant that in the spiral condition the dots towards the centre had a greater change in angle than those towards the periphery. This constant speed does not emulate true self-motion stimuli, but was recommended by Morrone et al. (2000) as it allows the self-motion compatible condition to be much better matched to the non-compatible stimulus. Each dot had a limited lifetime of between 100ms and 400ms. The stimulus was presented at a high enough frame rate to provide the illusion of motion. In the hemifield

conditions, the same density of dots was displayed, but only on the left or right third of the screen (2.42 degrees from the central fixation cross) thus displaying only about 300 dots. In the nine spiral condition, the spirals were tiled in a 3x3 grid, covering the whole screen. The motion went through a time-varying sequence of rotating-expanding-rotating-contracting with the amount of expansion and rotation changing continuously following orthogonal sine functions. These sine functions had a period of 3.14 seconds (so one full sequence of expansion-rotation-contraction would take this period to complete). This was based on the stimulus designed by (Wall & Smith, 2008), which was intended to maximise the BOLD response.

A second type of localiser was introduced after the first few participants, due to poor contrast between the two conditions. The 3x3 grid of 9 spirals was replaced by a random motion condition. In this condition the dots moved in straight lines in completely random directions, at the same speed as in the other localisers.

Distractor Task

A distractor task was used throughout the localiser. The fixation cross changed colour (blue, yellow, green, grey, orange) every two seconds. The participant was instructed to count how many times the cross turned blue. Every 75 seconds they were asked for this total with the message ‘Press the button if there were N blue crosses.‘, which appeared for five seconds. The value N was correct 60% of the time, one too high 20% and one too low 20% of the time. The participant responded by choosing whether to press the button on the button box and then restarted counting from zero.

4.3.6 Preprocessing Psychophysics data (behaviour)

Because the participants were required to wait until the end of each trial before responding, the response time was not considered of interest. The only relevant behavioural feature from the fMRI study was whether the response was correct or incorrect. This allowed us to determine whether a. the participant was fully attending the stimuli. b. whether the thresholds had been set correctly.

The behavioural adaptation-effect was calculated by considering the number of correct trials: subtracting this number for trials after those of the opposite direction, from those of the same direction, then dividing this by the total number of trials. This gives an adaptation-effect value for each participant for each trial type. Any adaptation would be indicated by a positive value.

4.3.7 Preprocessing fMRI data

fMRI data from both the functional localiser and the experiment were preprocessed using the SPM toolbox (SPM8, Wellcome Department of Imaging Neuroscience, London, UK). First they were realigned to the first volume of the time series, to minimise motion artefacts. Second the data was smoothed with a 3mm FWHM kernel. The images were kept in individual space, to avoid artefacts caused by spatial normalisation.

4.3.8 Localising ROIs

Functional Localiser

As mentioned earlier, Cardin & Smith (2010)'s research found the VIP contrast described detects another region (p2v) near VIP. To avoid misidentifying p2v for VIP, the clusters were chosen by comparing their normalised locations to the Talairach coordinates from both Cardin & Smith (2010) and Wall & Smith (2008). Similarly, for the hMT+ ROI, the coordinates of the clusters were compared with those from Dukelow et al. (2001) and Fischer et al. (2011). Huk et al. (2002), Smith et al. (2006) and Yan & Wu (2010) did not report cluster coordinates.

The five conditions (spiralling, random, static, left-hemifield, right-hemifield) became the 5 regressors (the blank was excluded, as it would be linearly inseparable from these five and the constant regressor). To identify VIP a contrast between the spiralling and the random motion conditions was used. To identify hMT+ the contrast between random motion and static dots was used. The voxels (presumably in left hMT) which don't respond to left-ipsilateral stimuli were identified by contrasting the left-hemifield stimulus with the full field spiralling stimulus (repeated similarly for the right-hemifield). Region hMST was expected to be the difference between the identified hMT+ complex and the hMT regions.

A later use was made of the unimodal auditory conditions in the experiment to allow the primary auditory cortex to be localised. A contrast was made between the auditory condition and the rest condition. A review by Vouloumanos et al. (2001) looked at the coordinates of A1 reported by several studies. Table 4.3, reproduced from their paper, illustrates the variability in the coordinates of A1. Similar variation was found across participants by Penhune et al. (1996) using anatomically defined landmarks (Table 4.4). Penhune et al. (1996) also provided a useful guide to anatomically defining A1. With a wide

Study [1]	Stimulus	Talairach Coordinates		
		x	y	z
Benson et al. 2000	music	-62	-23	16
Binder et al. 2000	tones	-52	-42	6
Celsis et al. 1999	tones	-30	-52	30
Mummery et al. 1999	noise	-54	-38	8
Zatorre et al. 1992	noise	-58	-21	8
Vouloumanos et al. (2001)	nonspeech	-63	-42	13

Table 4.3: Table of the coordinates of A1 from five studies cited by Vouloumanos et al. (2001).

range of coordinates from both anatomical and functional localisers there may have been a risk that A1 was misidentified. To gain assurance of correct localisation, the author contacted Pascal Belin (Centre for Cognitive Neuroimaging, University of Glasgow) an expert in auditory fMRI. He recommended that A1 is determined by the auditory contrast combined with the use of Heschl's gyrus as an anatomical

[1] See Vouloumanos et al. (2001) for these references and further detail

Axis	Range		
x	29	to	61
y	-35	to	-8
z	6	to	16

Table 4.4: Penhune et al. (1996) range of A1's location (Talairach coordinates).

landmark, and confirmed the regions for two of the participants ECN13 and SMW29 were correctly identified as A1.

Freesurfer (V1)

A V1 ROI was extracted from the anatomical image using Freesurfer (Fischl et al., 2002)[2]. The tool was used by running Oliver Hinds shell script (version date: 2007-06-16) which allowed the anatomical constraints described by Hinds et al. (2008) to be used.

Motivation for use of A1 instead of PT

Our experiment required the selection of a unimodal auditory region. Two obvious candidates stood out, the primary auditory cortex (A1) or the planum temporale (PT). In this experiment we chose to use A1, after originally starting the analysis with region PT. Below we've tried to summarise the motivations for this choice.

We briefly considered using normalised anatomical atlases to defined the ROIs. FSL (Jenkinson et al. (2012), FMRIB software library, University of Oxford, UK. version 3.1.8) provided a probabilistic map of the Planum Temporale. This was converted to an ROI for each subject in subject-space using SPM's normalisation function. Pascal Belin [personal communication] indicated that PT would not make a good ROI for the analysis, due to the ROIs large size and poor alignment with real anatomical features. So it was rejected in favour of the A1 localiser (described above). Secondly, the failure to detect significant changes in the BOLD response in the first few participants, in region PT, suggested the ROI would be a poor choice. This problem was repeated when DCM was tested, indicating that the mean time course of the region would not produce a sufficiently clear response to the stimuli.

For A1 to be a suitable substitute we needed to be convinced that it would represent similar information. Zvyagintsev et al. (2009) reports that A1 encodes auditory motion and Alink et al. (2011) also indicate that a trend exists for successful decoding of auditory direction from A1. Even if such auditory motion information does not directly exist in A1, the connectivity analysis does not require the feature of interest to be directly distinguishable, but just that it is represented within the region - this is almost certain for A1, which appears to be a critical region on the hierarchy of auditory motion processing. Auditory motion appears to be well represented in A1 in the electrophysiological literature, for example

[2]Download the freesurfer tool at http://surfer.nmr.mgh.harvard.edu

Ahissar et al. (1992) report that most of the neurons they recorded from in A1 were 'azimuth sensitive', that is, they were sensitive to the direction of the stimulus. Over a third were particularly responsive to the movement of the auditory stimuli.

4.3.9 Structural Equation Modelling

Introduction

Functional connectivity is defined as 'temporal correlations between spatially remote neurophysiological events', and is often investigated using structural equation modelling (SEM) (Friston et al., 1993). In contrast, DCM attempts to describe the connectivity between underlying physiological mechanisms. This has the advantage that it is investigating the influence of one region on another, can potentially reveal the direction of the effect and can use the temporal responses of those regions. To achieve this it has a series of generative models to link possible underlying neural activity with the functional BOLD response. These models are, necessarily, relatively complicated.

As an alternative to DCM, SEM provides a very simple method for looking at correlations in activity between regions. This simplicity comes at a cost: the direction of causality can not be determined, and the inference is about the functional connectivity (not the underlying effective connectivity). However, it is useful here as a complement to DCM. Note that the SEM used here is without latent variables (i.e. has no hidden regions), and so the computation being performed is equivalent to path analysis.

We originally attempted to use the tools provided by J. D. Steele (https://sites.google.com/site/fmrisem), linked from the SPM webpage, to develop a Reticular Action Model (RAM). Two problems were found with these tools; first, they didn't provide confidence intervals on the model parameters. Such confidence intervals could be calculated using the Hessian matrix, but it was not clear how accurate the Hessian would be. The second problem was that the predicted (model-generated) covariance matrix was often non-definite. Unlike the OpenMX tool, the matrix was not made using Cholesky composition (which ensures a positive definite matrix).

Instead we used the OpenMX software package (version 1.2.4-2063, http://openmx.psyc.virginia.edu, University of Virginia). The tutorial by Dorothy Bishop, albeit for twin-studies, provided a useful and quick introduction to OpenMX and R (http://dbtemp.blogspot.co.uk/).

Processing

The BOLD signal mean time courses for the three regions of interest were extracted using the MarsBaR SPM toolbox (version 0.42). The covariance matrix between the three regions was calculated (using MATLAB), separately for the two conditions: One covariance matrix for the condition in which the visual stimulus was more coherent, and another for the converse condition. The structural equation modelling was performed using the OpenMX toolbox, running in the R environment (version 2.15). The covariance matrix (from the MATLAB script) was read by the openMX R script, which then estimated the model parameters and their confidence intervals.

4.3.10 Dynamic Causal Modelling

Introduction

Dynamic Causal Modelling (DCM) is a method designed to investigate connectivity changes between different brain regions. One starts, generally, by defining several competing models of connectivity. These models consist of several brain regions, joined by directed connections. In this study the BOLD signal in these regions is known (from the fMRI data). A forward (generative) 'balloon' model is used to simulate and predict how neuronal activity within the regions generates the BOLD signal recorded in the fMRI study. Known, deterministic, experimental variables (i.e. values either controlled or observed during the experiment), can have two influences on the model. First, they can directly cause responses by modulating nodes of the model. For example, the presence or absence of a visual stimulus could be introduced to the model by modulating the neuronal response in an early visual area. Second, the experimental variables can influence the model by modulating the connections between regions. For example a variable representing the subject's attention could be used to modulate the connection between low and high level sensory processing regions.

The various models (potentially of different complexities) have their parameters estimated through a generative process. The likelihood of each model is computed, given the known fMRI data and the accuracy of the model output. The next stage is generally to compare the competing models using Bayesian Model Selection (BMS). BMS allows models of different complexities to be compared; although a model with more parameters will generally produce a closer approximation to the actual BOLD response observed, but this comes at the cost of extra complexity and parsimony. This added complexity is taken into account by BMS, which means, in general terms, that one needs stronger evidence to add more parameters.

A full description of DCM's internal models and parameters is beyond the scope of this thesis. Further details can be found in Friston et al. (2003).

Model Connectivity

The family of models of interest in our experiment connect a pair of (putatively) unimodal sensory regions (hMT+ and A1) with a (hypothesised) bimodal integration region (VIP). There are several hundred combinations of topologies and modulations which could describe the connections between these three regions. Prior to modelling, we shall select those models we believe are likely, given prior anatomical and function studies. The models being compared will also be chosen to test the hypothesis of interest: That the connectivity between the unimodal regions and the bimodal area is modulated by the coherence of the stimulus.

Section 4.2.3 outlines the known connections between these regions. Based on the known connectivity, it seems reasonable to compare the following models. The first (without any coherence modulated effect), consists of the three regions (hMT+, A1 and VIP), with reciprocal fixed connections between hMT+ and VIP, and between A1 and VIP (figure 4.11A). This model is, in effect, our null hypothesis:

There is no modulation in the connectivity between regions due to the coherence of the stimuli. The

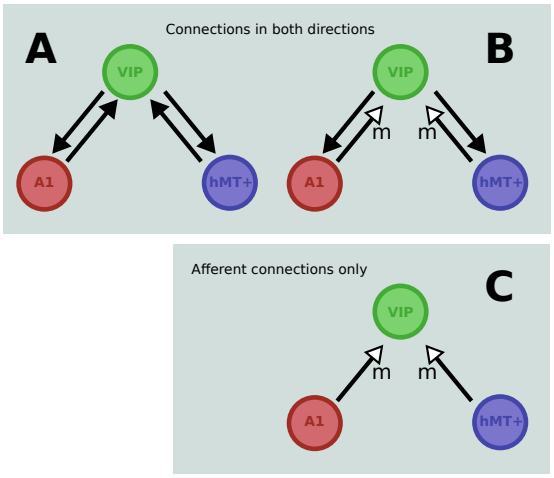


Figure 4.11: The three Dynamic Causal Models compared in this analysis. A and B are reciprocal (believed to be the most likely models), while C is feed-forward only. A comparison between B and C allows us to test which of these is most likely, while a test between A and B allows us to test our primary hypothesis that the connectivity is modulated by the coherence of the two stimuli.

second model (in figure 4.11B) is identical to the first model except for the addition of modulation on the afferent connections to VIP. The modulation by the coherence was specified by a regressor describing when the visual stimulus was more coherent than the auditory stimulus. Similarly the inputs to the two regions were both represented by the same regressor indicating the times when the stimulus was being presented to the participant. Figure 4.12 illustrates an example GLM used to help define the inputs to the DCM. The first regressor is a direct input to both sensory regions, while the second regressor modulates which of the sensory regions is most connected to the integration region (VIP).

Each region of interest consists of two cortical regions (one in each hemisphere). An analysis could also have been done with each area split into two regions. This would have offered far more prior choices regarding connectivity, and may have been useful in taking into account the rightward hemispheric bias, regularly reported in auditory motion processing (Schönwiesner et al., 2007). One might also argue that including these extra regions would fit the data better. However, the connectivity between hemispheres has considerable uncertainty. It is very common to combine the two hemispheres in studies of these areas, (e.g. Smith et al., 2007; Wolbers et al., 2011) and by combining the hemispheres, the time course of the two halves will still be reflected in the overall time-course for each region, making the model fairly immune to differences between hemispheres.

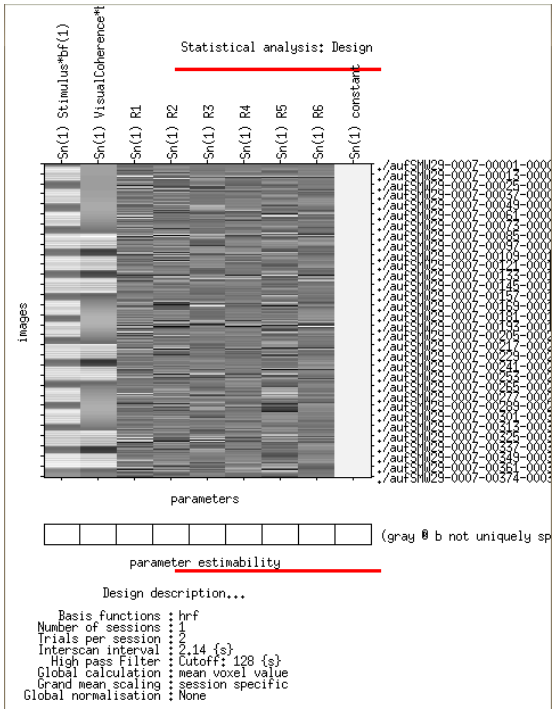


Figure 4.12: General Linear Model defined for Dynamic Causal Modelling. The regressors are, left-to-right, 1) stimulus presence, 2) coherence (set to one if the visual stimulus is more coherent), 3-8) motion regressors (from realignment), 9) constant term.

4.3.11 Decoding and correlation

Decoding

Instead of purely investigating the change in a region's average BOLD signal, the ensemble activity of the voxels within the region can be used to investigate a region's activity in more detail. The most common method is to attempt the 'decoding' of the region's activity using multivoxel pattern analysis. This involves training a classifier using a portion of the data, then testing the classifier on the remaining data. If the region's activity provides information about the regressor being classified, then the classifier should be able to predict the regressor's values with an accuracy greater than chance. The fMRI volumes were aligned as described earlier, but were not spatially smoothed.

For more details about decoding and classifiers refer to chapter 2. Below, the background and factors associated with the particular regions being tested in this chapter are summarised.

Decoding in VIP Zhang & Britten (2004) report that, in macaque, optic flow information is represented in a clustered manner, which could even be columnar. The results from chapter 2 suggest that, in spite of alternative explanations (e.g. maps Freeman et al., 2011), multivoxel decoding methods can be driven by column or clustered organisation. This suggests that there is a reasonable expectation that one

will be able to decode the location of the centre of expansion in VIP. Obviously such decoding depends on such organisation being conserved in humans, being large enough to shape voxel-scale responses and strong enough to provide a signal large enough to be detected.

Decoding in A1 Zvyagintsev et al. (2009) reported successful decoding of auditory motion in A1. Alink et al. (2011) expressed concerns about potential confounds associated with decoding direction of motion, but they themselves almost decode motion in A1 with significance ($p < 0.10$). The decoding reported is of extremely different auditory stimuli: A single sound source moving from left to right. In our stimulus, the cues are far more complex and the difference between the conditions is considerably more subtle: a slight change in the heading of the apparent selfmotion. Therefore we have considerably less confidence in the successful decoding in this region.

Decoding in hMT+ Although visual motion was decoded successfully from hMT+ in section 2, the difference between conditions in this experiment is far less dramatic, with most of the dots moving in a similar direction in the two conditions. For region hMT, simple motion is probably the most well represented stimulus. A small portion of the screen, near the fixation cross will experience considerable change between conditions, as dots will move down-and-left across the central visual field when the centre of expansion (COE) is to the right, and will move down-and-rightwards when the COE is on the left. This difference may be possible to decode, if the receptive fields of the region are small enough, and enough voxels cover cortex which represents this part of the visual field in the region's retinotopic map.

The other key subfield, MST, may respond to self-motion congruent cues, possibly with cells which respond to particular locations for the centre of expansion (Duffy & Wurtz, 1995). Therefore, the two directions of the visual stimuli may be decoded, as for our prediction for VIP. Again, this also depends on a sufficient response size and clustering.

Correlation method

As a complement to the standard decoding method described, we also used a multivoxel pattern analysis described by Haxby et al. (2001) and used more recently by Morgan et al. (2011). This method involves splitting the data into two parts, A and B. The pattern was considered correctly classified if the correlation of the two patterns was higher between the same condition in the two groups than between opposite conditions across the two groups. The difference between these correlations was used as a metric of pattern decoding accuracy.

The data was split originally by session, but this seemed unreasonably challenging. Subject movement and scanner-drift between sessions makes voxel responses vary between sessions. Instead the sessions were split into parts which were combined. Note that the blocks were kept intact (i.e. no block was split between groups), this meant that the autocorrelation caused by the HDR delay would not give any confounding bias.

4.3.12 Adaptation

Introduction

The term ‘waterfall effect’ was coined by Thompson (1880) quoting R. Addams from 1834, ‘after looking for some time at a waterfall and then at the waterworn-rocks immediately contiguous, he saw the rocky surface as if in motion upwards’. This after-image is now believed to be due to neural adaptation; the reduction in a neuron’s firing rate over time, when exposed to a constant stimulus. The psychophysics result of such adaptation are described in section 3.4.9. For this chapter the key features of adaptation aren’t the behavioural or perceptual effects, but are the expected effects on the fMRI BOLD activity being recorded. Many fMRI studies rely on the adaptation effect to investigate neural activity (Krekelberg et al., 2006). In summary, the BOLD response to motion in a second trial is reduced if the motion is in the same direction in the first trial. Similar adaptation effects are reported in several other visual domains (e.g. orientation Tootell et al., 1998). The study by Smith et al. (2007) found adaptation to auditory moving stimuli (by direction) in the planum temporale, but they also found such adaptation to stationary stimuli. The stimuli used in that study were single simple sound sources moving from the left to the right or *vice versa*, or sounds from single points. The current study used far more complex, multi-source cues, with the position and location of sounds designed to minimise confounds caused by simple effects such as the change in sound intensity.

Method

The adaptation analysis was performed by creating a design-matrix for SPM, in which each session had two key regressors: one for those trials in which the previous trial’s stimulus was of the same condition, and one regressor for those trials in which the previous trial had the opposite condition. Figure 4.13 illustrates the design-matrix used by SPM to test for adaptation.

Once defined and estimated, a contrast between the regressor for different previous stimuli minus the same previous stimuli was calculated. This contrast value and its associated p-value was reported, for each participant. The contrast values were combined in a t-test across participants to test for group-level effects.

Type of Stimulus

Before testing the method on the stimulus directions, the methods were validated by applying them to the more trivial sort of adaptation caused by the presence or absence of the stimulus. Adaptation was tested for in hMT+, comparing visual trials which follow a previous visual trial and those visual trials which did not (this therefore includes trials with auditory stimuli and rest trials).

We tested for the adaptation in the cortical region believed to respond to the stimulus. So adaptation to the visual stimulus direction was tested in regions hMT+ and VIP, while adaptation to the auditory stimulus direction was tested in regions A1 and VIP.

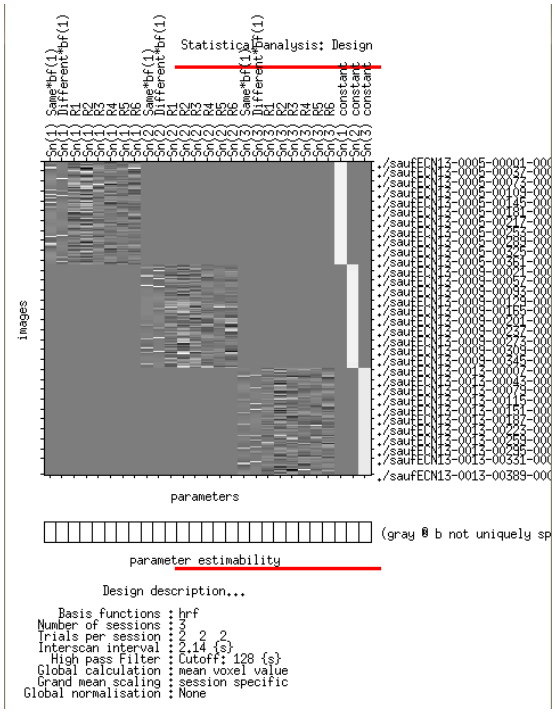


Figure 4.13: The design matrix for the adaptation analysis

4.3.13 BOLD response to coherence

The BOLD signal values were calculated by finding the mean of all volumes within each block. The regressors were shifted by 2 volumes to reflect the haemodynamic response function. Although possibly not the most accurate or complete measure of the BOLD response, it allows a reasonable comparison between regions to get an idea of the signal strength and reliability. Using marsbar, the BOLD time-course for each region was extracted, giving a BOLD response value for each region, for each stimulus. These were compared to see which stimuli produced a greater response. These differences were compared across subjects using t-tests to get group-level statistics on the effects of interest.

4.4 Results

4.4.1 Behaviour

As mentioned in section 4.3.6, the only behavioural feature of interest was the accuracy of the participants.

Unimodal thresholds

In the low coherence group, the unimodal part of the main experiment used the 4 up 1 down (84.1%) thresholds, while in the high-coherence group 100% coherence stimuli were used.

Low Coherence Group In the low-coherence group, the participants' accuracy in the unimodal sessions averaged 85.8% for visual stimuli and 83.7% for auditory stimuli (these are both within the 95% confidence interval 81.9%-86.3%). The variation in unimodal accuracy between subjects appears to be within the 95% confidence intervals of the expected accuracy for both stimuli (figure 4.14A,B). This suggests the training and coherence selection methods employed are robust and provide good estimates even in an MRI scanner environment.

Participants KHS22 and MHD12 inadvertently had their left and right audio signals swapped. In the unimodal condition this had no effect (as it merely resulted in the sounds being in the opposite direction to that generated by the software). In the bimodal conditions the reversed sounds caused a conflict condition - and so those participants were unable to be used in the bimodal analysis later.

High Coherence Group For visual stimuli the five participants had a combined accuracy of 98%. Individually they all reached high levels of accuracy, with the least successful participant still failing fewer than 1 in 16 trials. This suggests the visual stimulus is easily distinguished at 100% coherence.

The auditory stimuli proved to be more challenging at 100% coherence, with an average accuracy of 92.4% across participants. The variation between participants was also quite noticeable with the least successful only achieving 88% accuracy, while the most successful reached 98%.

This suggests that in future research the auditory stimuli may need to be made easier. Some of the difficulty could be due to the efforts made to reduce simple-cues around the auditory motion (e.g. which side the sounds begin or end on).

Bimodal thresholds

Low Coherence Group In the condition in which the auditory stimulus has a higher coherence (set to the 84% threshold vs 71% for the visual stimulus) the participants achieve 82.9% accuracy. The predicted accuracy (based on Bayesian integration) is 87.2%. It might be that integration is not occurring in this condition, however, looking at individual participants (figure 4.14C,D) it appears only one is significantly below this accuracy value. The small number of trials makes it difficult to judge whether integration is occurring in this case, however the earlier work in chapter 3 suggests that these stimuli are successfully integrated by the participants.

Due to an error in the scanner setup, two of the participants were given conflicting stimuli (the left and right audio channels were reversed). In this case we would expect 73.1% accuracy (towards the higher coherence stimuli, see equation 3.26 for derivation). In practice the accuracies were 81.0% and 92.9% (auditory>visual) and 40.5% and 17.9% (visual>auditory). Oddly this appears to indicate

that the auditory stimulus captures the participant's preferred direction. The small number of trials and participants in this 'accidental' experiment makes it difficult to draw any definitive conclusions however.

High coherence Group One would expect the participants to do better in the bimodal condition than in the unimodal condition. In this group, one stimulus was at 100% coherence, while the other was at the 84.1% accuracy threshold. In the condition with a greater visual coherence the average accuracy was 91.4%, while in the converse condition it was at 96.7%. Figure 4.14E,F illustrates the individual results across participants. Although not at 100% the average number of mistakes was less than one in every ten trials.

Behavioural Adaptation

As in the psychophysics experiment in chapter 3, the response of participants can show signs of adaptation. The high-coherence group were given 100% coherence stimuli during the fMRI sessions, so the adaptation analysis focused on the low-coherence group.

Figure 4.15A illustrates the adaptation effect, across just the five (non-conflict) low-coherence subjects. There appears to be less variance in the bimodal conditions across participants, and they may be appear to experience a greater adaptation effect (although not significant, one-way ANOVA across session types, p = 0.70).

The high-coherence group did have adaptation in the bimodal condition, even though the coherence of one of the stimuli was always at 100% (figure 4.15B). Comparing between the unimodal and bimodal conditions is problematic as the design of the sessions was different, with the participant potentially anticipating that the bimodal stimuli are in blocks, each containing 3 trials of each direction. The bimodal conditions were significantly more influenced by the adaptation effect (paired t-test between the means of the bimodal values vs the means of the unimodal values, p = 0.0038, t = 6.04, N = 5, 4 dof). It is likely this effect is partly due to participants expecting equal numbers of the two trials types within each block.

Combining the two groups gives an even greater significance to the difference in adaptation between the unimodal and bimodal conditions, with the bimodal conditions having a greater adaptation effect (with p = 0.0017, 4.415, N = 10, 9 dof).

4.4.2 Localisers

hMT+

The region most easily localised using the functional localiser session was the hMT+ complex (which includes both MT and MST). As described in the methods, the region was localised by both the stimulus contrast and by reference to several previous papers identifying MT/MST/hMT+. Detailed descriptions of the localising are described in appendix C. In summary, ten of the twelve participants whose data was used in further analyse had region hMT+ localised with thresholds at least $p < 0.05$ (FWE), of the

two remaining participants the thresholds were at least $p < 0.01$ (uncorrected). For most participants, the clusters identified appeared to lie within the predicted bounds of previous published studies.

MT/MST

The hMT+ complex consists of at least two regions (MT and MST) involved in visual motion processing. To distinguish between these two regions, advantage was taken of MSTs receptive fields which are purported to extend into the ipsilateral hemifield. Upon analysing the data for several participants however, we found no evidence for a cluster in the region of hMT+ in the ipsilateral hemisphere (while the contralateral hemisphere still had a large BOLD response, see figure 4.16). This result was apparent across subjects, leaving us unable to distinguish the two regions. Although not critical for future analysis, this was unexpected. As insufficient participants had MST localised, the whole of hMT+ would be used as a single region of interest for further analysis.

Full details of the localisation results across participants can be found in appendix A.2.1.

Primary Auditory Cortex (A1)

As described in the methods section, the primary auditory cortex (A1) was not originally planned as a region of interest. However, the planum temporale provided a poor region to use for several reasons and so A1 was chosen instead. Although no specific localiser was available to find A1, it was realised that the unimodal sessions could act as a localiser. The region defined could then be used for analysing the bimodal sessions' results.

The region was successfully localised using a whole brain threshold of $p < 0.05$ FWE for eight of the participants, while four had one or both hemispheres identified with an uncorrected threshold with a p value of, at most, 0.005. Detailed A1 localisation results are available in appendix A.2.2.

Ventral Intraparietal Cortex

The first few participants (SMW29, KUE24, JBA10, EKW30 and LCA24) were given the original localiser, based on the comparison of a self-motion compatible stimulus contrasted with nine small 'spirals', described in section A.1.1. This was found to perform very poorly, with the region not being identified at all in four of the five subjects.

The region expected to contain VIP was found to have a strong response (compared to stationary dots), suggesting that the reason for the poor response was due to the control stimulus being too similar to the self-motion stimulus. One hypothesis put forward was that the 3x3 grid of spirals was a poor choice of stimuli because the central tile in the grid would provide a congruent self-motion cue across 4.7° of the visual field, covering the whole area of foveal vision.

One possibility was to replace the grid with a 4x4 grid, causing the central region to be split into four different spirals. It may be that this would not avoid the problem sufficiently. To reiterate the purpose of the control; we needed a stimulus which contain motion at the same speeds and dot-density as the

congruent self-motion stimulus, but did not contain self-motion congruent features. The most obvious stimulus was random-dot motion.

The 3x3 spirals had been chosen to ensure the control was as similar as possible to the congruent condition. In contrast, this alternative stimulus may be too distinct. There was concern that the clusters the new contrast identified may include regions beyond VIP. It was felt that this could be mitigated by careful comparison of the cluster coordinates with those previously reported and careful inspection of the anatomy in the region of each cluster.

Two of the five former participants (EKW30 and SMW29) were re-recruited to allow the new localiser to be used. One of these (SMW29) underwent the whole experiment again, using the high-contrast stimuli. We were unable to rescan participant LCA24. As a fall-back we found that the VIP region (and probably surrounding parts of the Intraparietal sulcus) could be approximately identified by a small subset of data from the unimodal contrast (visual stimulus vs rest).

The remaining twelve participants were all tested with the new localiser, allowing putative VIP to be identified. Appendix A.2.3 describes the localisation in detail. In general a more generous threshold was required for VIP localisation than for the A1 or hMT+ regions, with $p < 0.05$ (uncorrected) being the most liberal threshold.

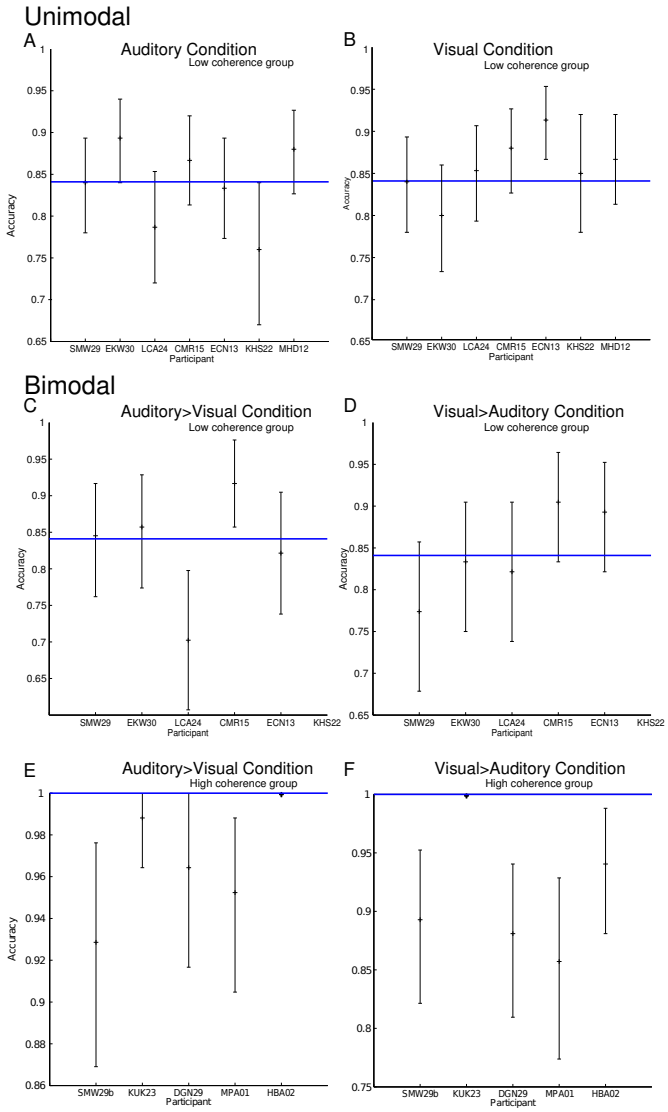


Figure 4.14: Behavioural data from the experimental sessions in the fMRI scanner. A and B, the low-coherence group's unimodal condition; C-F, the bimodal condition. The errorbars show 95%-confidence intervals discretely sampled (in steps of 1/150 for the unimodal and 1/84 for the bimodal conditions) using the binomial distribution. The blue lines indicate the target accuracy in that condition.

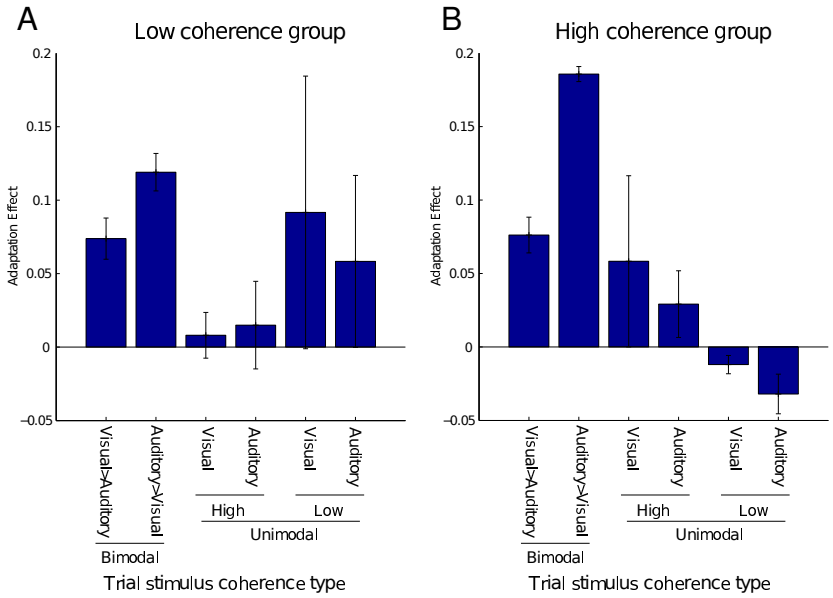


Figure 4.15: Behavioural Adaptation: A Low-coherence group (excluding conflict subjects). B High-coherence group. Error bars represent one standard error across subjects.

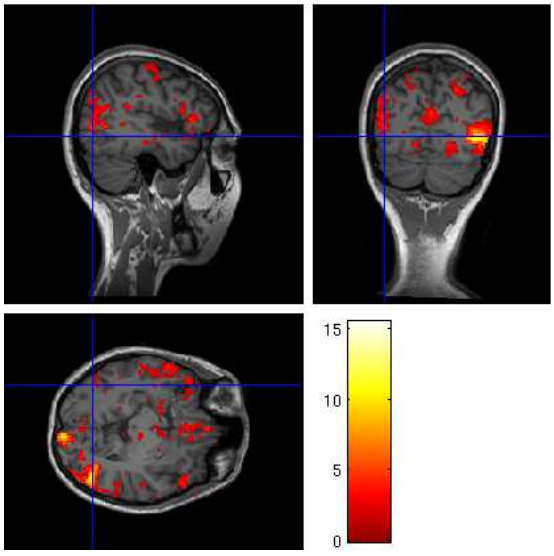


Figure 4.16: Example showing the absence of an ipsilateral BOLD response to visual motion in one visual hemifield. Crosshairs approximately mark the expected location of MST. Colourbar indicates t-statistic values (Threshold $p < 0.05$, uncorrected). Image generated by the SPM toolbox. Subject MPA01.

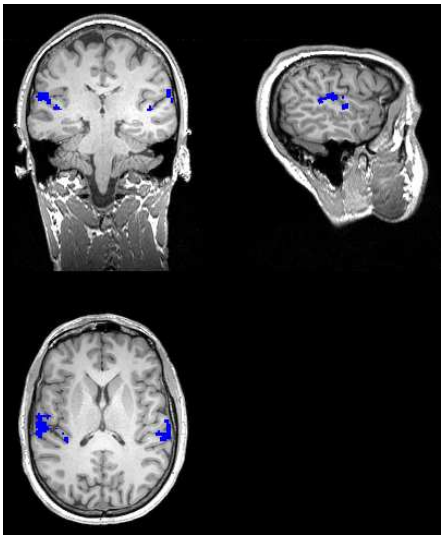


Figure 4.17: Example of A1 localisation (participant ECN13).

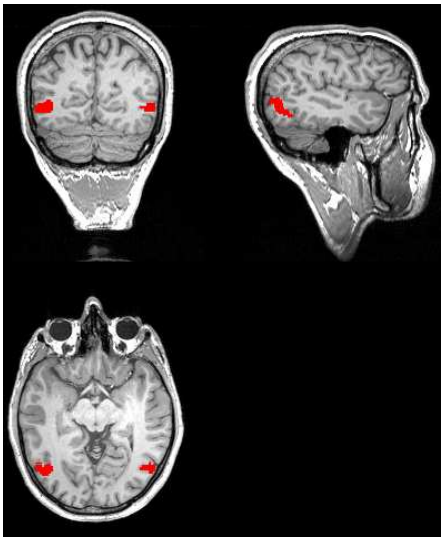


Figure 4.18: Example of hMT+ localisation (participant ECN13).

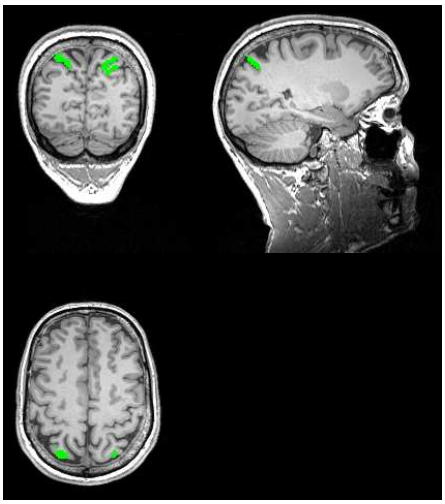


Figure 4.19: Example of VIP localisation (participant ECN13).

4.4.3 BOLD Response

Before the full analysis, we summarise the BOLD signal from the different regions. These results are interesting, first to see how strong the BOLD response was and second, whether there was any sign of superadditivity. The experiment was not designed to test this hypothesis, and the differences in stimuli and sequence order between unimodal and bimodal sessions meant the comparison would be difficult, but worth qualitatively considering.

hMT+

Earlier Study Results The region we expected to most clearly show a response to the stimulus was hMT+. For comparison, the BOLD responses in the earlier fMRI experiment (see chapter 2) have also been calculated. In the stimulus condition it was found the BOLD response was, on average 1.73% greater than in the rest condition (s.d. 0.57% across participants). Figure 4.20 illustrates this strong response, with an increase of at least 0.75% in all subjects.

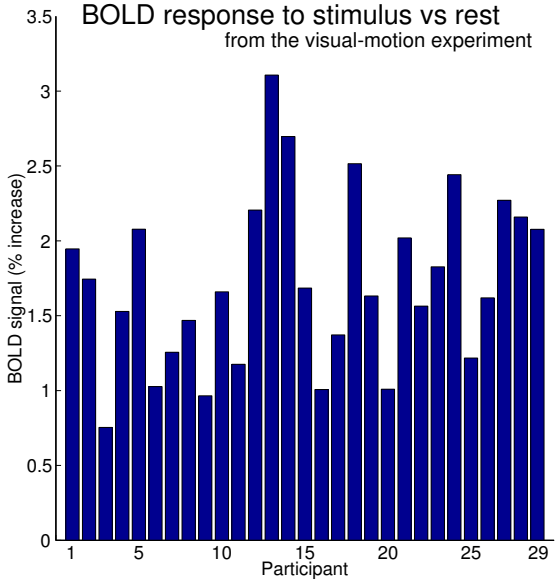


Figure 4.20: BOLD response in hMT+, to visual motion stimulus (vs blank screen) across participants for earlier study (chapter 2).

This study's results In contrast to the strong, reliable BOLD signal in the earlier study, the signal from this study was much less reliable (figure 4.21A). This is probably mainly due to the event related design (necessary for the investigation of adaptation). The method of BOLD signal extraction was also bias towards the block design. The mean increase was only 0.24% (s.d. 0.31%). The increase was significant across subjects, but not as much as one might expect ($p = 0.0217$, $t = 2.67$. 11 dof). Clearly the within subject GLMs are far more significant, so this result is not a dismissal of the stimulus. This

result is useful to illustrate the small signal strength in this experiment.

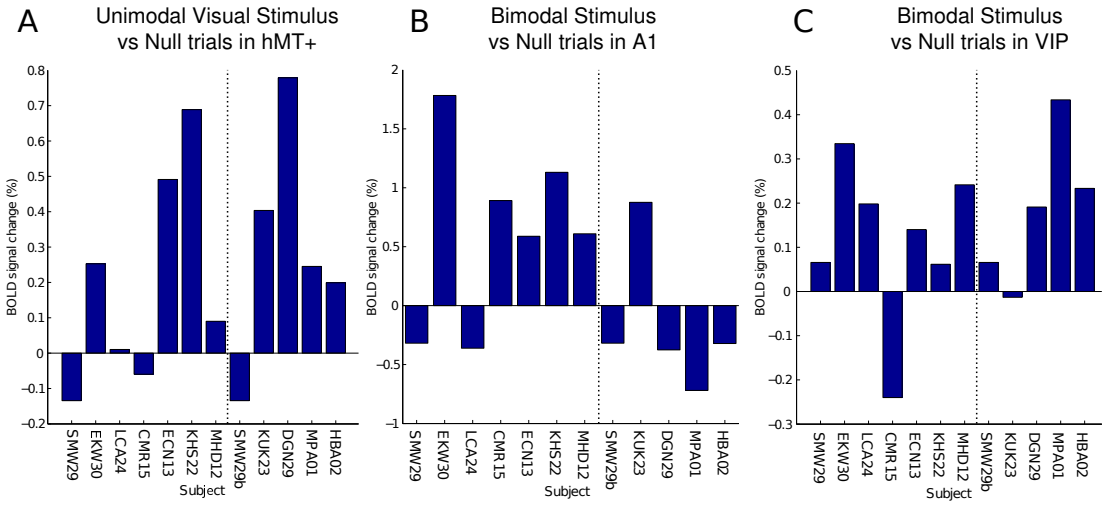


Figure 4.21: Effect of stimuli on BOLD response compared to the rest periods. A, Unimodal visual stimulus vs null trials in hMT+. B, Bimodal stimulus vs null in A1. C, Bimodal stimulus vs null in VIP. Note y-axes vary between subfigures.

A1

The region A1 was defined using the unimodal sessions as a localiser. We cannot therefore look at the BOLD response in the unimodal sessions (as this would be a form of circular analysis). Instead, the bimodal sessions were used, with those conditions with a stimulus (which will include both visual and auditory stimuli) are compared to those conditions with the blank screen.

Figure 4.21B shows that this contrast appears even weaker than in region hMT+, with a similar mean (0.29%) but a greater standard deviation (0.79%). This leads to the increase not being significant across participants (p=0.2140, t=1.319).

There appeared to be a slight bias towards a larger BOLD response in those participants given a lower coherence stimulus. A two-sample t-test does not find the difference to be significant (p=0.086, t=1.90, two-tailed), although with such small sample sizes it is hard to draw any conclusions.

VIP

Compared to these two unisensory regions, VIP appears to have a slightly more reliable BOLD response to the stimuli. Figure 4.21C illustrates VIP's BOLD responses to the bimodal stimuli (vs the rest blocks). This reliability may be chance, as the actual BOLD response increase is, on average, only 0.14%. It is the low variance between participants (s.d. 0.17%) which explains the significance (p=0.016, t=2.83, one-tailed). This reliability may be due to the activity of the region, but is equally likely to be to do with the accuracy of localisation.

Super-additivity

Although not originally a hypothesis for this experiment, the effect of superadditivity was investigated in VIP, but was not detected. Figure 4.22 illustrate the response to the two unimodal stimuli and the bimodal stimulus. The superadditivity hypothesis suggests that the response to the bimodal stimulus should be greater than the sum of the two unimodal stimuli. This clearly was not the case in the majority of participants, or in their average BOLD response values.

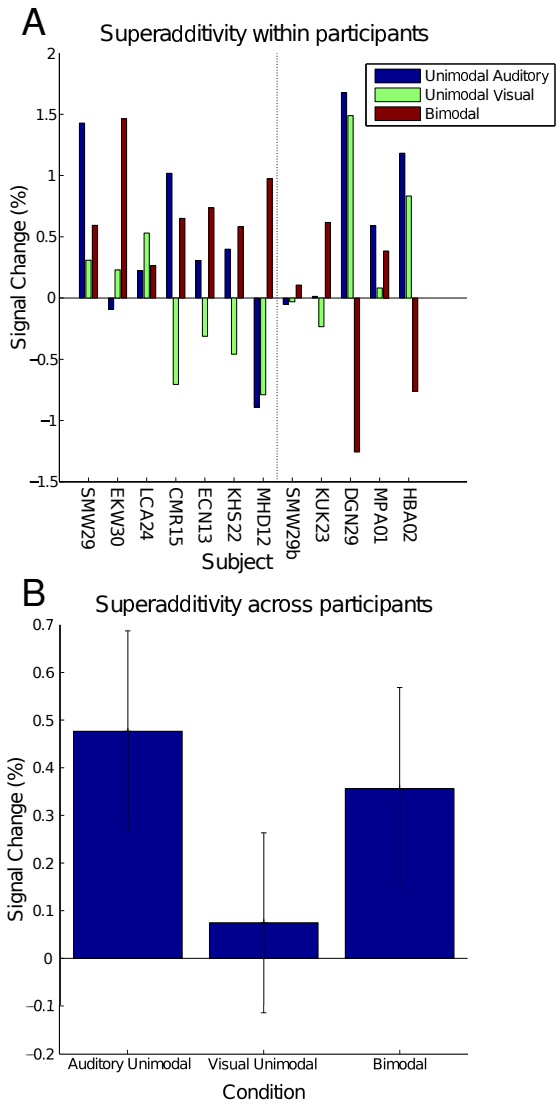


Figure 4.22: Superadditivity: BOLD signal in VIP in the two unimodal and the bimodal condition. No evidence for superadditivity (A) within or (B) across participants. Error bars indicate 95% confidence intervals.

Summary

In summary, the BOLD responses were quite weak (compared to the earlier) study. This may be due to the event related design, but may also be due to the stimuli, which are less coherent, lower contrast, and have lower speeds and dot-densities. The stimuli do seem to induce BOLD signal changes, but it is worth bearing in mind the low signal strength and the effect of this on the following analyses.

4.4.4 MVPA Decoding

Stimulus Type

Before attempting to decode more interesting features, we tested the decoding algorithm by attempting to decode the presence of the stimulus in the unimodal condition.

As expected, the classifier was able to successfully classify the condition for all four regions tested (Table 4.5).

Region	Across Participants				Notes
	Mean Accuracy	Standard Deviation	p-value	t-value	
V1	90.72	4.04	$p < 0.001$	35.00	Visual Trials
MT	86.80	7.62	$p < 0.001$	16.76	Visual Trials
A1	83.72	6.98	$p < 0.001$	16.71	Auditory Trials The A1 region was defined by the stimulus being classified, thus this result is compromised.
A1	67.42	5.37	$p < 0.001$	11.24	This used the bimodal stimuli sessions to provide the signal to decode.
VIP	65.01	10.77	$p < 0.001$	4.80	Bimodal stimulus

Table 4.5: Classification of stimulus presence.

There appeared to be greater decoding accuracy in the low coherence group, but this was initially confounded as more time was available for scanning the earlier participants and so most of the low-coherence group had more unimodal sessions than the high-coherence group. To see if there were differences between the two groups (and so whether the coherence of the unimodal stimulus had any effect), decoding was repeated using just the first two unimodal sessions (allowing the two groups to be compared, Table 4.6). Two sample t-tests were performed for MT, A1 (unimodal) and VIP). Decoding in MT was significantly greater in the low-coherence group (p=0.029, t=2.15, one-tailed) but not when corrected for these multiple comparisons.

The low-coherence group had region hMT+ identified before the high-coherence group. It is plausible that my methods around threshold and cluster selection changed over time. To mitigate this potential

confound, the same comparison was repeated decoding the stimulus type using the activity in V1, a region localised anatomically using an automated software tool. This region was found to also have a greater decoding accuracy in the low-coherence group (p=0.031, t=2.11, one-tailed).

Region	Mean Decoding Accuracy		one-tailed p-value	t-value	Notes
	Low Coh Grp	High Coh Grp			
V1	92.14	86.40	0.031	2.11	
MT	88.86	80.80	0.029	2.15	
A1	82.85	84.60	0.6896	-0.4112	The A1 region was defined by the stimulus being classified, thus this result is compromised.
A1	68.85	65.40	0.2923	1.1117	This used the bimodal stimuli sessions to provide the signal to decode.
VIP	69.00	59.40	0.1358	1.6224	

Table 4.6: The Classification of stimulus presence between groups (unimodal stimuli).

Stimulus Direction

Obviously of more interest is whether the direction of the stimulus motion can be decoded in any of the three regions. It was expected that 100% coherent visual motion would be decoded in hMT+, but it appears this was very limited.

Table 4.7 presents the classification scores across subjects for the stimuli and regions tested.

Stimulus	Region	Average score	p-value
visual	hMT+	51.80	0.06
visual	VIP	52.16	0.08
visual	A1	53.49	0.038*
visual	V1	54.14	0.048*
auditory	VIP	46.97	0.92
auditory	A1	51.30	0.25

Table 4.7: The Classification of stimulus direction (p-value one-tailed, across participants).

It is worth mentioning that this poor decoding of direction information was the motivation for the switch to a higher coherence in the remaining five participants.

It was noted that occasionally very low classification scores were recorded. For example, the classify data from hMT+ in subject DGN29 by visual motion direction, was decoded with only 37% accuracy

(p=0.0060, not multiple-comparisons corrected). Although this may have occurred by chance, it can be an indicator of over-fitting. Such a problem is generally resolved by introducing additional data. To achieve this, three adjustments was made to the decoding algorithm. First, to increase the number of volumes provided to the classifier, the blocks were not averaged, instead each volume was given separately to the classifier. Second, in chapter 2 V1 was reported to have a higher decoding accuracy than hMT+, so the decoding was applied to V1 instead. Third, the Least-Angle Regression (LARS) classifier was used instead of SVM. LARS is ℓ_1 regularised, which means the coefficients for most of the variables will be set to zero, effectively restricting the input to a small number of voxels. This may protect it slightly from over fitting, and intuitively is a good match to the data from V1, in which most voxels will probably not be providing useful direction information. These three changes appeared to improve classification accuracy to an extent, providing a significant classification score across the group for decoding visual direction in V1 (p = 0.016, t = 2.45, N = 12, one-sided t-test). Figure 4.23 illustrates the decoding accuracies across participants. The accuracy is still quite poor, although one participant's classification reached 68% correct.

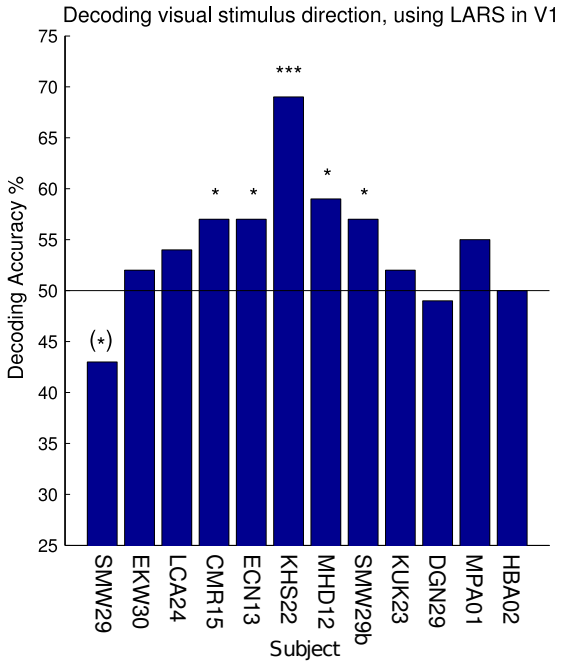


Figure 4.23: Decoding the direction of visual unimodal stimuli from V1 using the LARS classifier. Significance determined through permutation testing (and checked against the binomial distribution). Bracketed star indicates significantly below chance. Results not corrected for multiple comparisons.

Correlation

Decoding simple tasks To confirm the correlation classification method was correctly implemented, the tool was first tested on an easier task. In this case we decided to test it on the data from the experiment

described in chapter 2. The groups being compared were leftward and rightward trials.

Figure 4.24 shows the results from this analysis. Although the result is significant across subjects ($p=7.1660 \times 10^{-7}$, t=6.1894, df=31) it doesn't appear to have done as well as classification using a SVM; the MVPA classifier in contrast achieved above chance decoding on each subject, individually.

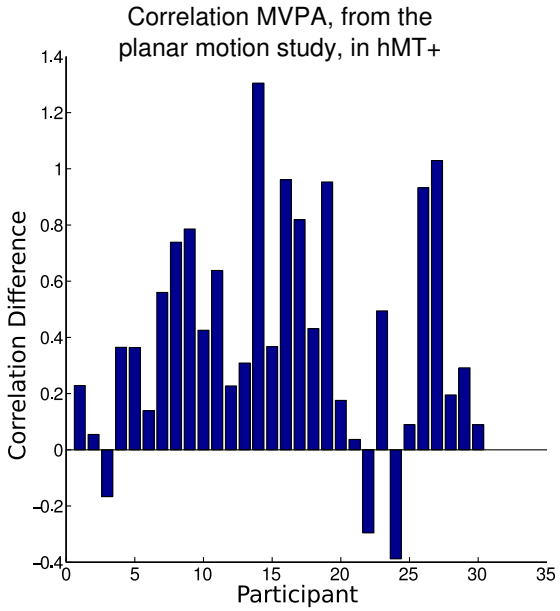


Figure 4.24: The difference between the correlation of trials of the same direction and the correlation of trials of opposing direction [from the MT experiment], across two groups of trials. A positive value indicates the classifier is finding the trials of the same direction to be more similar to those of opposite directions. Region of interest: hMT+.

To test this tool further, this study's unimodal stimulus type (i.e. visual vs auditory) was used as the parameter being decoded, and the tool was applied to data from hMT+ (figure 4.25). The correlation between stimuli of the same type is significantly greater, across subjects than the correlation between stimuli of the opposite type (p=0.00091, t=4.4971, df=11). This confirms that our regressors and tool are functioning correctly.

Decoding Stimulus direction using adaptation analysis As with the standard classifier, decoding stimulus direction appears not to be successful: There were no significant results in any region. The only exception was in A1, classifying the unimodal direction-of-motion regressor. However, this region of interest was defined by the unimodal stimuli and so this data is being sampled twice, and so this result is potentially compromised (p=0.018 t=2.388).

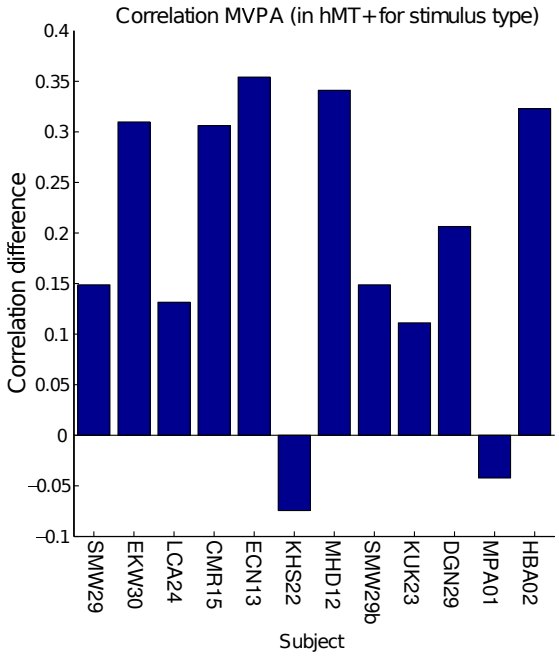


Figure 4.25: The difference between the correlation of trials of the same type (e.g. both visual) minus the correlation of trials of opposing types (e.g. visual and auditory), across two groups of trials. A positive value indicates the classifier is finding the trials of the same type to be more similar to those of opposite directions. Region of interest: hMT+.

4.4.5 Dynamic Causal Modelling

Model Shape

Before comparing the models of interest (with regards to our hypothesis), we briefly compared the results of the modulated model (figure 4.11B) with that of the same model without efferent connections (figure 4.11C). This was to test whether the DCM results agreed with our hypotheses surrounding the reciprocal connections between the different regions.

Figure 4.26 illustrates the results of comparing these models. In the high-coherence group especially it appears the model containing reciprocal connections is considerably more supported by the data than the model without such connections. There appears to be a division between the two groups, with the higher coherence group separating the models significantly more (averaged session pairs, compared groups using the wilcoxon-rank test, $p = 0.048$).

Influence of Coherence

As described in section 4.3.10 the models to test our main hypothesis differed purely by whether the afferent connections were modulated by the coherence regressor. Figure 4.27 illustrates the results across subjects. The figure suggests that the model with connections modulated by the stimulus coherence is

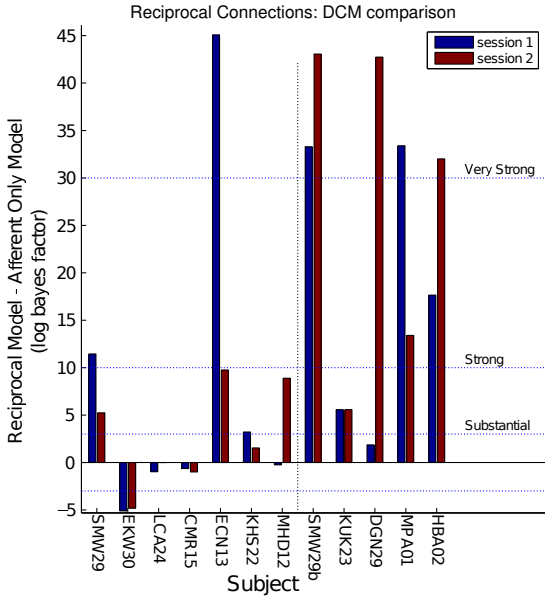


Figure 4.26: Comparing the modulated model with another model without reciprocal connections. The y-axis indicates the log of the bayes factor between the two models (a greater number indicates the data was more likely given the reciprocal model). Subjects 1-7 were given the lower-coherence stimuli, 8-12 the higher-coherence stimuli. The two sessions were analysed separately.

more likely to have created the data than the model with unmodulated connections. Within the higher-coherence group, seven of the ten sessions have substantial or strong evidence towards the modulated model. The lower-coherence group seemed to have much weaker results (this difference was not significantly different).

Group level statistics To perform group-level inference on the model structure, we must choose how to combine the individual subject data. Stephan et al. (2009) described several ways of combining data from a group of subjects. For Bayesian Model Selection (BMS) one can, in general, either used Fixed Effects (FFX) or Random Effects (RFX) Bayesian model selection (BMS). The fixed effects model makes the (potentially invalid) assumption that the optimal model is identical across subjects and a group Bayes factor is calculated for the whole group. The method is also vulnerable to outliers (Stephan et al., 2010). In a random effects comparison, a different model is fitted for each subject, this avoids the problems associated with FFX, but is generally less statistically efficient (Kasess et al., 2010). A third option is to use the model evidences in a frequentist statistical test. This is particularly vulnerable to outliers causing it to lose statistical power.

FFX The FFX analysis was applied to the two groups separately (to avoid inhomogeneity). For the high coherence group the modulated model's log-likelihood was 141.4 greater than the unmodulated model. This is classed as very-strong evidence. The low-coherence group, in contrast, found that the

modulated model's log-likelihood was only 15.4 greater than the unmodulated one. This difference is apparent in figure 4.27. The difference in the coherence and usefulness of the stimuli is much greater in the high-coherence group; the 100% coherence stimulus is far more useful and perceptible than the 84% accuracy stimuli. This difference may explain the success of DCM in the high-coherence condition, compared to the low-coherence group.

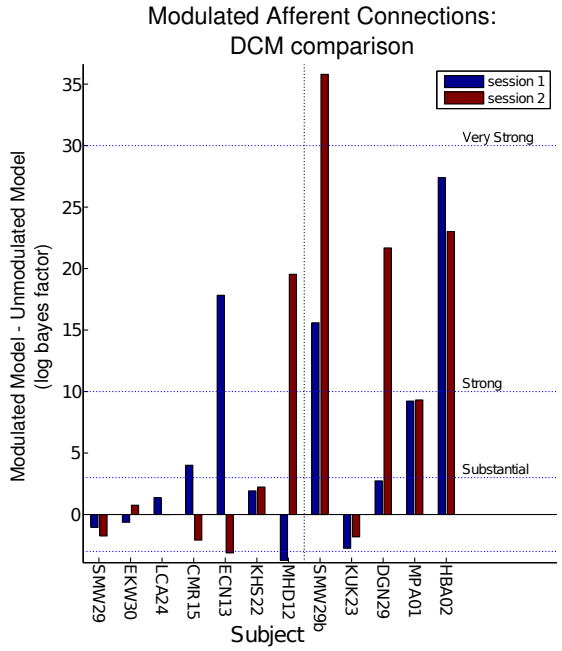


Figure 4.27: DCM results across subjects, showing the preference for the model with the modulated afferent connections over the model with the unmodulated connections. The sessions from each participant are analysed separately. Note that subjects 6 and 7 experienced the conflict condition in which the sounds directions were reversed. Participants 8-12 experienced the high-coherence stimulus set.

RFX The use participants make of different components of the stimuli may vary. This was indicated qualitatively by the variety of responses participants gave in post-experiment interviews. This variation may cause the assumptions behind the FFX analysis to be compromised, given that participants may attend to different features within the stimuli. To incorporate this variation and allow all participants data to be combined, an RFX test was applied to all participants. Figure 4.28 illustrates these comparisons. The RFX comparison combining all but the two participants who experienced conflicting stimuli had an exceedance probability of 0.0596. With these two subjects included the exceedance probability dropped to 0.0181. This suggests our model is significantly more likely than the unmodulated equivalent.

Model Parameters Curious as to whether the modulation was in the direction we expected, we looked at the parameter estimates provided by the DCM. As the high-coherence group appeared to have the most

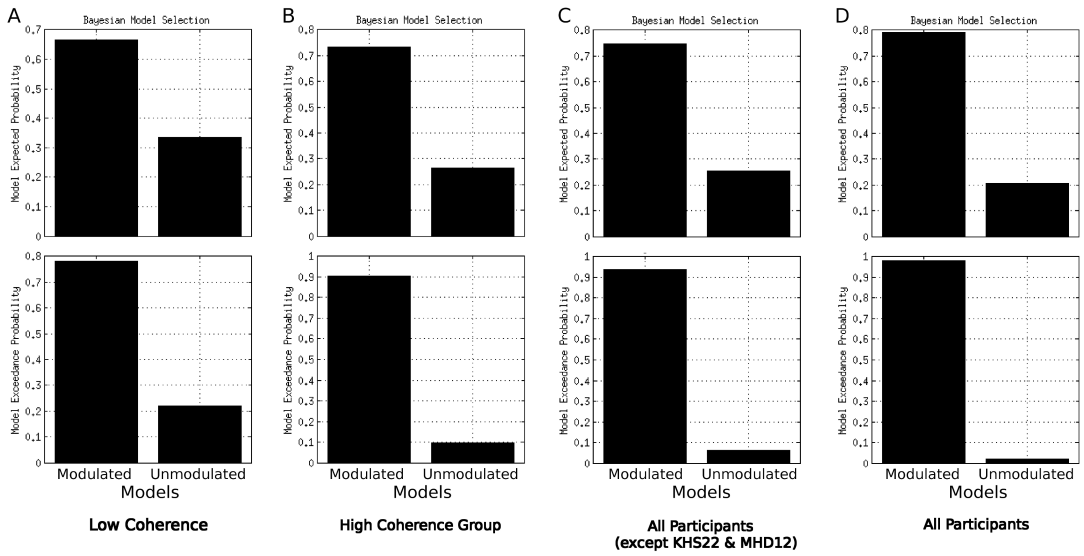


Figure 4.28: DCM results combined using a RFX model, showing the preference for the model with the modulated afferent connections over the model with unmodulated connections. Note that D includes the two participants who experienced the conflict condition in which the sounds directions were reversed.

significant results, we concentrated on their results. Figure 4.29 shows how the coherence regressor influences the connectivity between the two input regions (hMT+ and A1) and the integration region (VIP).

The results suggest our hypothesis about the connectivity is highly doubtful. We would have expected the coupling between the visual area (hMT+) and VIP to be greatest when the visual stimulus is most coherent. This would be indicated by the blue bars in the figure being positive. Similarly we would expect the auditory region (A1) to be most coupled with VIP when the auditory stimulus is most coherent, which would have been indicated by negative yellow bars.

Goodness of Fit

The fit of the BOLD signal's time course for each of the three areas does not have a clear scale-invariant metric in DCM. However, the R^2 coefficient of determination, as used by Lohmann et al. (2012), is a potentially useful approximation. The values of which are reported in table 4.8. There is a significant difference in the goodness of fit between regions (one-way ANOVA, p = 0.0015, F = 7.97), with VIP being significantly less than A1 or hMT+ (separate t-tests, p = 6.1×10^{-4}, 0.0018, t = 4.7435, 4.088, respectively).

Conclusion

The introduction of modulated connections between regions appears to be more likely, even when the extra model complexity is taken into account (as is standard in DCM). However, the direction of the

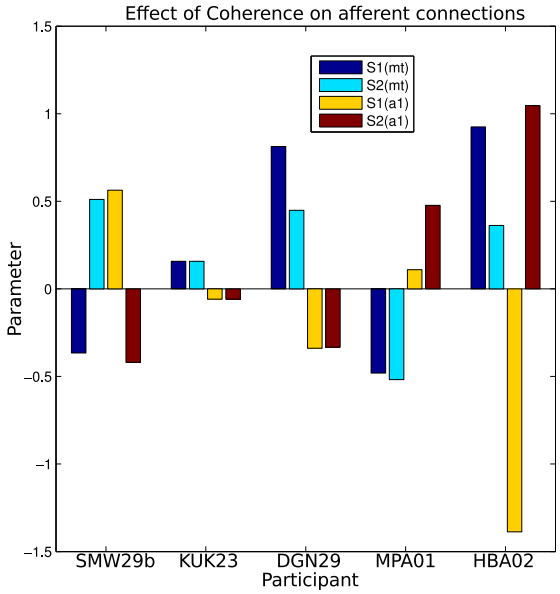


Figure 4.29: Model parameters estimated by DCM. The blue bars show how the input from the visual area (hMT+) is modulated by the coherence-regressor, a positive value indicates that the connection is positively coupled when the visual stimulus is more coherent than the auditory stimulus. Conversely, negative values indicate the activity in hMT+ is negatively coupled with VIP. The same applies to the yellow bars, with positive values indicating that the activity in A1 is positively coupled with VIP when the visual stimulus has a greater coherence. One would expect the blue bars to be positive and the yellow bars to be negative, reflecting the expected effect of visual coherence on the coupling between regions.

effect is not consistent across subjects. This means the FFX analysis is probably also unreliable.

4.4.6 Structural Equation Modelling

To supplement the DCM analysis, we used SEM to test for changes in functional connectivity. To analyse the data, we estimated the parameters of the SEM in the two conditions separately. The value of the regression coefficients between the input regions (hMT+ and A1) and the integration region (VIP) were then compared between conditions. Figure 4.30 shows how the parameters change between conditions. We would hypothesise that the connection from the visual area (in red) would be greater in the high-visual coherence condition, and so be positive in the figure. We would also expect the connection from A1 to VIP to be the converse.

The connection from hMT+ to VIP does trend ever so slightly towards being greater in the high visual-coherence condition, compared to the connection from A1 to VIP (mean difference in increase of coefficient between the two regions: 0.0298, std: 0.0973. t-test: p=0.31. t=1.06), but is far from significant. One can not therefore draw any conclusions about the influence of the coherence on the connectivity between these three regions.

Region	A1	hMT+	VIP
Minimum	0.018	0.009	-0.001
Maximum	0.251	0.372	0.069
Mean	0.099	0.144	0.021
Standard Deviation	0.061	0.115	0.022

Table 4.8: Values of R^2, coefficient of determination, for each region, across subjects (in session 2 only).

Figure 4.31 attempts to display the change in connectivity between conditions in a potentially more intuitive manner. We would expect (given our hypothesis) that the plotted points would appear in the top-left of the the figure (with A1's connection to VIP reducing with the increase in visual coherence, and hMT+'s connection to VIP increasing, simultaneously).

4.4.7 Adaptation

Type of Stimulus

As with the earlier analyses, to confirm the software tools and methods were functioning, we tested them on merely the presence of the stimulus, not its direction, with the expectation that adaptation would occur if the prior trial consisted of the same stimulus (compared to rest). Across participants the effect was highly significant, when looking at the response to the visual trials in hMT+ (p = 0.0012, t = 4.3042, paired t-test across subjects. Figure 4.32A). Similarly for the auditory stimulus in A1, significant adaptation was found both within many participants and across subjects ($p = 5.34 \times 10^{-}04$, t = 4.82, paired t-test across subjects. Figure 4.32B). In VIP the adaptation effect was less clear, although still significant; using the visual stimulus (p = 0.025, t = 2.21. mean = 0.79%, std = 1.23%, figure 4.32C). The lower significance appears to be more to do with the large variation between participants, which is probably due to variations in the accuracy of the localiser. Finally, the response difference in VIP to visual trials following auditory compared to visual trials following the null trial was compared (figure 4.35). Not quite significant across all participants (p=0.077, t=1.52), two individual participants reach significance (after bonferroni multiple comparison correction), and taking just the high-coherence group finds a strong significance (p=0.0092, t=3.85).

Stimulus Direction

The unimodal and bimodal directions of motion were hypothesised to produce adaptation. For example a unimodal leftward visual trial was expected to respond less strongly after a similar trial, compared to its response after a rightward visual motion trial.

The fMRI results for the unimodal and bimodal sessions from hMT+ did not show significant evidence for adaptation (for example, in the bimodal experiment, the adaptation between directions had p=0.9141, t=-0.1103, across subject, paired t-test.) and none of the participants results were significant

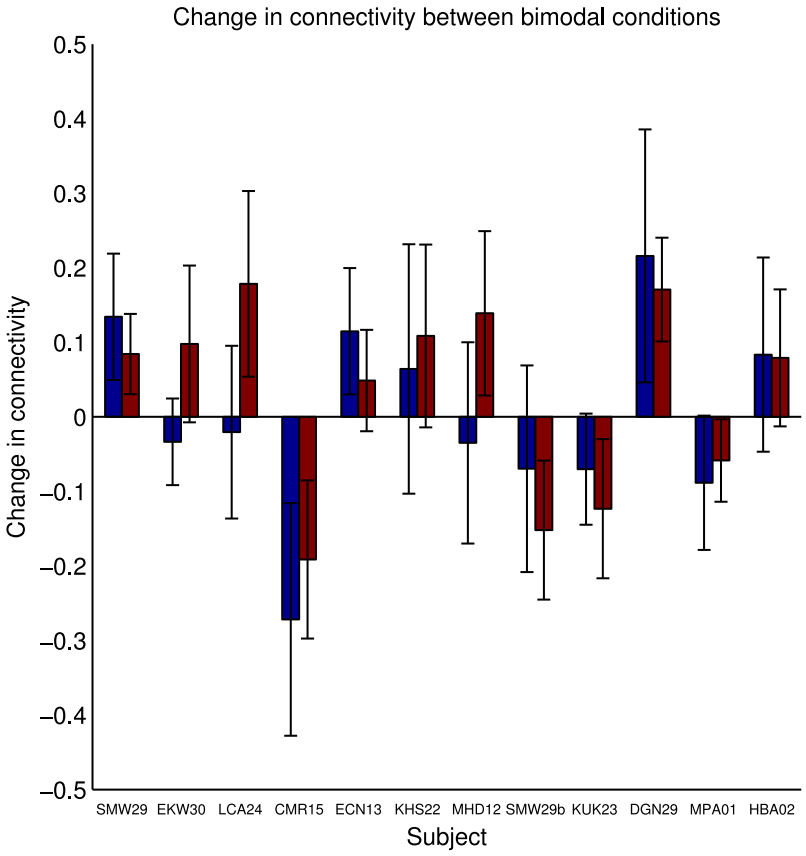


Figure 4.30: SEM results: The difference in the regression coefficients between the two conditions (high visual coherence condition minus high auditory coherence condition). Blue: Change in A1 to VIP coefficient, Red: hMT+ to VIP coefficient. Error bars indicate one standard error in the estimate of each parameter.

(once corrected for multiple comparisons). Similarly the fMRI results from A1 did not show significant evidence for adaptation either as a group (p= 0.9716, t=0.0364) or within subject (once corrected for multiple comparisons). It should be noted that within the low-coherence group the auditory adaptation contrast was found to be significantly above chance (mean=1.15, std=0.92, p=0.016, t=3.31, N=7), note that A1's ROI is defined by this stimulus and so these results are possibly confounded.

Cross Modal Adaptation

The original experimental design envisioned testing for cross-modal adaptation due to the direction of the stimulus. For example, if the previous unimodal trial was leftward auditory motion, we hypothesised that we would find adaptation effects in VIP, if the next trial was also leftward but of the visual stimulus. The absence of adaptation within modalities suggests cross-modal adaptation will not be found.

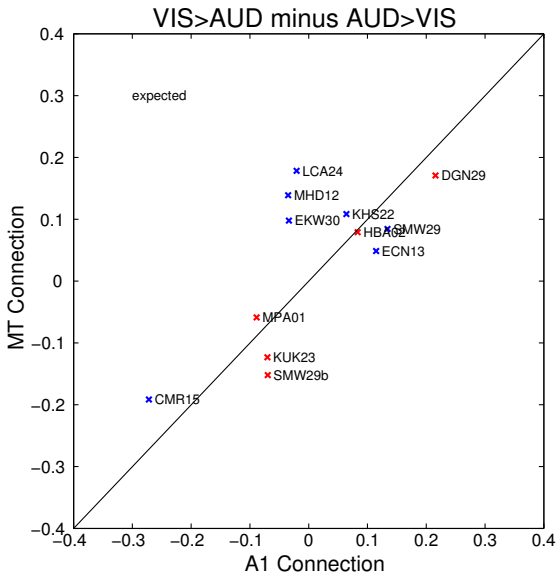


Figure 4.31: SEM results: The change in connectivity given a higher visual coherence. The hypothesis would imply the points would lie towards the top-left of the figure.

4.4.8 Other Relevant Regions

Several other related cortical regions are known of, which may have relevance to this study. These regions are briefly described and their responses to the stimuli reported below.

V6

The dorsomedial cortical visual area (V6) was first identified by Allman & Kass (1975), in the Owl Monkey. It is found in primates at the most dorsal extreme of the extrastriate cortex, bordering the parieto-occipital sulcus (see appendix A). The region is certainly associated with motion-stimuli, and some fMRI studies suggest it is associated with self-motion congruent stimuli Pitzalis et al. (2010).

The region was localised using the same contrast as used for VIP; self-motion stimuli *vs* random-dot motion. An ROI-based t-test across participant *beta*-responses was used to test for the region's response to the visual stimulus compared to the null-trials, and the auditory stimuli compared to the null trials. The region had a significant increase in the BOLD response across subjects for the visual condition (t, 5.9538; p, 0.00014; df, 10), but not to the auditory motion condition (t, -0.7709; p, 0.4586; df, 10).

A classifier testing for patterns associated with the direction of visual motion was applied to the data from this region. Across subjects there was no significance above chance (accuracy, 49.27%; p, 0.748; N, 11). To confirm the region does have activity associated with the visual stimulus the type of stimulus was classified, and found to reach an accuracy of well above chance (p¡0.00003 in all subjects).

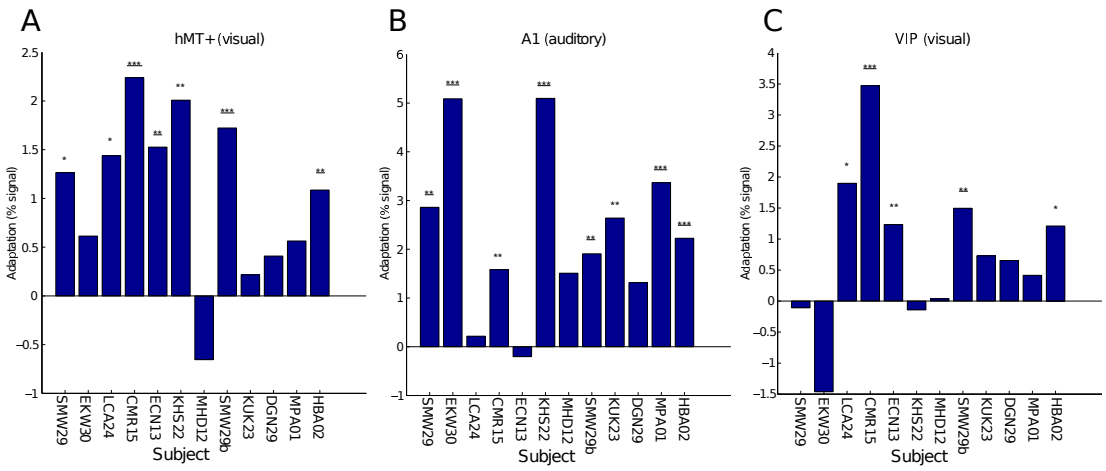


Figure 4.32: Adaptation by stimulus type. A, hMT+ (visual stimulus); B, A1 (auditory stimulus); C, VIP (visual stimulus). Asterisks indicate significance at the 5%, 1% or 0.1% level, prior to multiple comparisons correction. Underlined indicates they remain significant at least at the 5% level, after bonferroni multiple comparisons correction.

CSv

Another relevant brain region is the recently identified visual-response region, discovered by Wall & Smith (2008) and named the Cingulate Sulcus Visual Area. The region has not previously been associated with optic flow, but as described in more detail in section 5.1.4, the nearby posterior cingulate and the associated retrosplenial cortex are strongly associated with spatial memory, in particular the pathway via which visual information (e.g. optic flow) enters the head-direction and spatial processing system (Vann et al., 2009). More details of its localisation is described in appendix A.

As with V6, the region was localised using the same contrast as used for VIP; An identical ROI-based t-test across participant *beta*-responses was used. As with V6, the region had a significant increase in the BOLD response across subjects for the visual condition (t, 4.8978; p, 0.00085; df, 9), but not to the auditory motion condition (t, -1.2372; p, 0.2473; df, 9).

Classification again did not approach significance for either auditory or visual motion direction (accuracy, 48% and 50% respectively, p values, 0.25 and 0.95).

STS

The superior temporal sulcus divides the superior and middle temporal gyri, in the temporal lobe. It is likely that the region contains several, functionally distinct, areas. In macaque, for instance, a particular portion of the floor of this sulcus has been found through cell-recordings to contain neurons responsive to multisensory stimuli (Beauchamp, 2005). Called the temporal-parietal-occipital area (TPO), this region is likely to be roughly equivalent to the posterior portion of human STS (Beauchamp, 2005). Research has not yet reached the point of reliably distinguishing these subfields using fMRI, so this section will

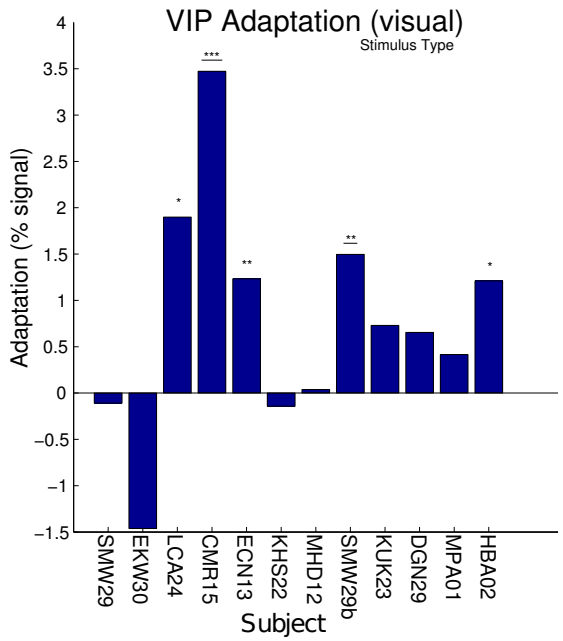


Figure 4.33: Adaptation: Comparing the response in VIP to visual stimuli following either auditory trials or trials with no stimulus. Symbols as in figure 4.32

follow the lead of Beauchamp (2005) and consider the whole STS region. The region has had several tasks and purposes assigned, including the perception of biological motion (Beauchamp et al., 2002) and the perception of others' gaze direction (Hoffman & Haxby, 2000).

We localised the region anatomically using FSL (Jenkinson et al. (2012), FMRIB software library, University of Oxford, UK. version 3.1.8), which provided an ROI for the whole STS.

The same ROI-based t-test across participant *beta*-responses was used to analyse the fMRI data. Unlike the V6 and CSv ROIs, the region had no significant increase in the BOLD response across subjects for the visual condition (t, 1.87; p, 0.10; df, 7) or the auditory motion condition (t, 1.02; p, 0.34; df, 7).

Only 8 participants were used for this region, due to problems with anatomical alignment. The weak results reported here my also be due to the poor localisation of the relevant area. In particular the STS is a large region, and the average of the whole ROI may mask any effect from a small restricted area within the ROI.

Man et al. (2012) found that audiovisual cross-modal classification performed significantly better in the posterior STS region than in the other regions they investigated. The stimuli used were individual objects, and not visual motion stimuli. However, the location of the pSTS, between visual processing regions such as MT and auditory processing areas, suggests it is a likely location for the integration of audiovisual cues, as indicated by Man et al. (2012).

Further inferences about the response of region STS are limited by the poor localisation. Contrasts were considered to localise the region using the data already collected. None of the contrasts available

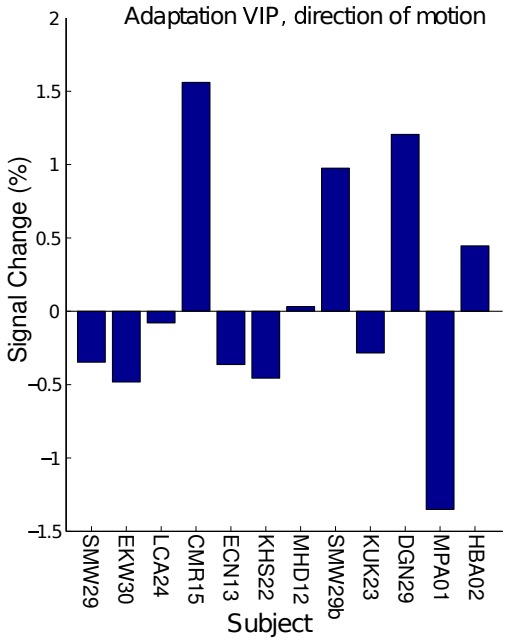


Figure 4.34: Adaptation results: Visual direction in VIP

seemed to identify the region at the location described in Man et al. (2012). Its importance for visual motion must be somewhat limited as it does not appear to strongly respond in particular to planar or selfmotion stimuli.

Summary

Although regions CSv and V6 appear to respond to visual self-motion cues (based on the successful localisation) they do not appear to have a strong preference for auditory stimuli. This suggests that they are primarily associated with visual processing and are not integration areas. The results for the STS region are far more ambiguous, due to the poor localisation of the region.

4.4.9 Eye Tracking Results

During the fMRI imaging, eye tracking was performed for many of the participants, to test whether they were successfully fixating, or if eye-movements were potentially confounding the BOLD signal. Figure 4.36 illustrates the raw density maps around the central fixation point for five of the participants. The standard deviation in the coordinates was compared between conditions (visual stimulus vs none) to test if people move their eyes further or more greatly in one condition than the other. The standard deviation was slightly larger in the stimulus condition (63 vs 57 pixels). No significant difference was apparent across participants (t-test across five participants, t=-1.28, p=0.27). Clearly only five participants is a small sample to test this, but for two of the five participants the no-stimulus condition had a greater standard deviation. In summary, there does not appear to be a strong increase in eye-movement in the

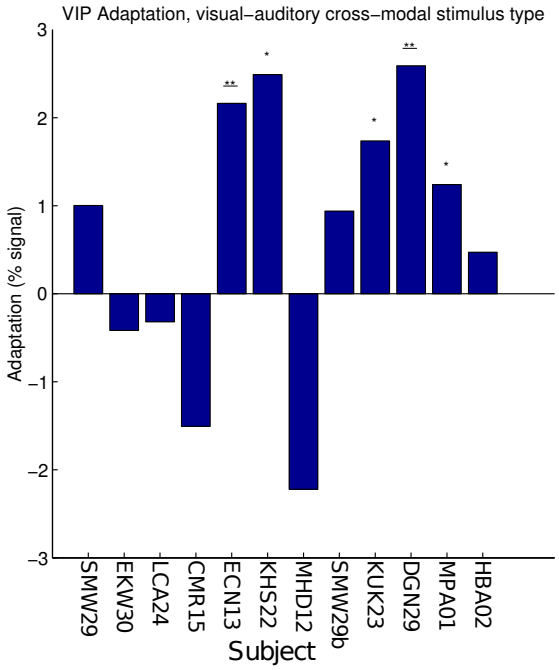


Figure 4.35: Adaptation to stimulus types: VIP BOLD signal reduced in visual trials following auditory trials. Symbols as in figure 4.32.

visual stimulus condition compared to the no-visual stimulus condition.

4.5 Discussion

4.5.1 Summary

In this chapter we have developed the self-motion stimulus and tested its utility in an fMRI study. The hypothesis that coherence controls connectivity was tested using DCM and SEM. The identity of VIP as an integrating region was also investigated, using both superadditivity and cross-modal adaptation. Most results were negative, suggesting that either the choices of stimuli, sample size, scan sequence or analysis were poor, or the effects of interest were not large enough to be detectable using these methods. Suggestions were made for improvements in localisers and stimuli, to allow future experiments to build on this work.

Behaviour

In terms of behavioural response the stimulus has been shown to be robust across participants, with all but one understanding the task and achieving the expected response accuracies in both the unimodal and bimodal conditions.

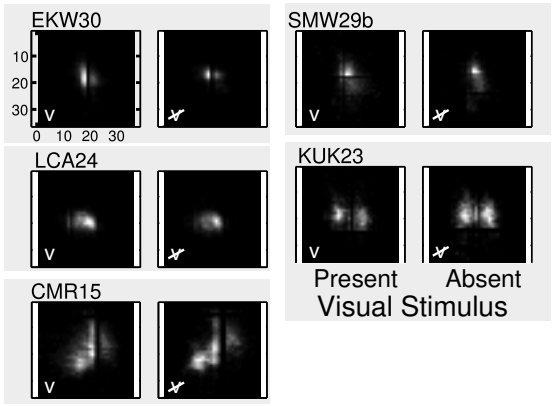


Figure 4.36: Raw eye track data: Density maps around the central fixation points for five participants. Blank areas caused by eye-tracker failing, where pupil was hidden or combined with reflection. Distances are number of pixels across screen.

The psychophysics adaptation results from the previous chapter were replicated in this study. The additional finding that the bimodal condition experienced greater adaptation was also noted.

Localisers and BOLD response

The regions of interest were localised using several different contrasts. Although the 9-spiral contrast had been reported to successfully localise VIP in previous studies, in this experiment it was found to be very poor across several participants. After replicating this poor detection level across several subjects, we replaced this stimulus with the random-motion alternative, which was found to be more successful. However, this localiser needs to be validated to confirm other regions are not being included in the contrast. The primary auditory cortex (A1) was localised post-hoc by using the unimodal sessions as a localiser. This provided a very robust localiser for A1, probably due to the considerable number of trials.

The BOLD response to the stimuli was less than reported in our earlier study. This is probably due to the event-related design and lower stimulus-coherence.

Finally hMT and hMST couldn't be reliably separated. The reason may be the short length of the localiser and so relatively weak signal. The localiser in this experiment consisted of only three minutes for each condition, giving 6 minutes for the contrast of interest. In comparison the sessions conducted by Dukelow et al. (2001) lasted at least 15 minutes for each participant, in Huk et al. (2002) the localiser lasted 25 to 50 minutes. However the more recent study by Wall & Smith (2008) only lasted 8 minutes, suggesting that our 6 minutes should provide some evidence of activity.

MST neurons have been reported to have receptive fields of between about 15° and 40° (Desimone & Ungerleider, 1986). It may be that the small size of the MST localising stimulus (only 6° high and 14° across) covered an insufficient area to cause cells in MST to respond.

Decoding

The stimulus type was decoded successfully in all regions. An unexpected but intriguing difference was found to exist in the decoding accuracy between groups, with the low-coherence group apparently decoding more accurately. With the tiny sample size, this result probably is not highly significant. Far more difficult was the decoding of the stimulus direction, this did not succeed in any region using the standard MVPA methods except in V1. The correlation-method of decoding only succeeded in A1, which was confounded by the method of its localisation.

Connectivity

DCM Initially we used DCM to analyse the connectivity of the three regions. The model containing reciprocal connections was found to be more likely than one without, confirming our prior beliefs based on anatomical literature. DCM appeared to indicate the connectivity was modulated by the balance of coherence of stimuli, however, further investigation of the model's parameters found such an inference is not necessarily correct, as the parameter estimates were not stable across participants, but rather varied between both positive and negative values. This instability may be due to differences in individual processing, attention or strategy but is also likely to be due to the methods used.

It may be that poor model fit is a key problem in this analysis. No separate analysis was performed looking at the success of the model fit. This decision was based on the work of Stephan et al. (2010), in which they write 'model fit does not need to be considered explicitly', although they do suggest that BMS is performed across all plausible models - one might argue the set of three models used in this paper (although based on anatomical evidence) is not sufficiently broad to achieve the level of model fit Stephan et al. (2010) intended. However, they later write that 'we cannot obtain false inference simply because of 'poor' model fit'(Stephan et al., 2010). In other words, our significant RFX result can not have been due to poor model fit. This assertion is questioned in a recent paper by Lohmann et al. (2012), in which the difference in the relative goodness of fit to the different regions is considered of particular concern. In this analysis the three regions did have significant differences in their goodness of fit, but it was unclear if the difference was large enough to cause the sort of problems highlighted by Lohmann et al. (2012).

SEM To try to analyse the same question with a different technique, structural equation modelling (SEM) was used to estimate correlations between the input regions and VIP in the two conditions. The hypothesis implied that VIP and hMT+ would be more correlated in the high visual coherence condition, and that A1 and VIP would be more correlated in the converse condition. However, no significant difference was found between conditions.

Adaptation

Changes in stimulus type caused detectable levels of adaptation in all three regions, but adaptation was not significantly revealed by changes in direction in any of the regions.

Of particularly interest was cross modal adaptation in VIP. Such adaptation would suggest that the same neurons were responding in both the unimodal auditory and unimodal visual stimulus conditions, in VIP. The experimental design was for this effect to be found by comparing the adaptation to direction between the two stimuli, but without direction adaptation within-condition, the equivalent bimodal comparison was not possible (figure 4.34).

Considering just the adaptation between modality types, it appears there is a significant effect of adaptation in region VIP to the two stimuli; with adaptation being found to occur in visual trials following auditory trials (compared to those following a null trial). However, it is not necessarily the spatial component of the stimuli which is causing the adaptation. For example it may be the additional cognitive load, attention or general increase in the average neural activity in the region which is causing the adaptation in the following trial. This result does add some weight to the hypothesis that there is an increase in activity in the same neurons for either stimulus type, implying that the cross-modal colocalised BOLD signals found in VIP are also, at least partly, found in the same neurons.

4.5.2 Analysis

Does coherence modulate connectivity?

Linear Summation Recently Pouget, DeAnglelis and Angelaki bolstered their argument that some form of the linear summation model drives integration (Fetsch et al., 2012). They recorded from cells in macaque MST and compared their responses to conditions in which visual and auditory motion stimuli were presented with different levels of coherence. Individual cells appeared to combine the two cues in a manner similar to that reported for the behavioural results of the animal. For example, at low-coherence, the vestibular stimulus is weighted more strongly in behavioural responses than the bayes-optimal model would predict. This effect is also found in the responses of individual neurons. To simplify somewhat, they suggest that the neurons' responses are approximately scaled by the coherence of the stimulus. They argue that this response scaling, combined with poisson noise (as they mentioned in their previous paper, Gu et al., 2008) matches the recorded integration results.

If the linear summation model is true, how can the results of earlier fMRI studies finding connectivity changes (e.g. Kreifelts et al., 2007; Nath & Beauchamp, 2011; Beauchamp et al., 2010) be explained? Kreifelts et al. (2007) reported the activity of the connected regions increased with increasing coherence, an effect which could confound the correlation result. Beauchamp et al. (2010) reported that increasing noise (decreasing coherence) leads to both an increase in activity in the unimodal regions and a decrease in the correlation with the integration region. They argue this is due to a reduction in the connectivity and possibly protects against the criticism levelled at Kreifelts et al. (2007). However, an alternative explanation would be that the extra activity in the unimodal region is filtered before it reaches the integration area. For example adding *extra* random-moving dots to our visual stimulus to decrease coherence would probably lead to an increase in activity in region hMT+, but not VIP (due to its preference for coherent motion stimuli). The correlation between the areas would be likely to fall as the two regions respond to different components of the stimulus (hMT+ to both the coherently and randomly moving dots, VIP to

the minority of dots just moving coherently). Therefore it is not the connectivity changing, it is simply that the response of the two regions is to different dimensions of the stimulus.

Controlling for such additional cues is a challenge for future experimental stimulus designs. In our experiment for example one can see that in retrospect the auditory stimuli should have been degraded in a different way, instead of the addition of extra noise. This would have made it much homologous to the reduced coherence in the visual field. During the development of the stimulus an attempt was made to degrade the auditory cue's motion by modifying the paths the auditory cues took. The small number of cues meant that motion was generally still perceived, just not in the direction intended, rather than a general reduction in motion perception. These issues led us to develop a stimulus in which the coherence was reduced by the addition of noise, but clearly a different method would be more ideal.

Future experiments could investigate these possibilities by altering coherence in two ways: Either by adding noise, or reducing the stability of the stimulus, one could test whether the changes in correlation were due to the addition of extra noise to the sensory regions (which is then filtered before reaching integration areas), or by the remove of coherence (which the model would predict would lead to a reduction in connectivity between regions).

A potential alternative method for analysing such data in the future could involve making use of this uncertainty. If the increase in the unimodal region's activity is assumed to be noise, could the effect this has on correlation be estimated? For example, a model could estimate the response to different stimuli, including white noise and the original signal in the various regions being compared. One could then ask if the reduction in correlation is fully explained by the change in response to coherence differences.

Returning briefly to the results of the DCM analysis; the RFX DCM analysis results did report a significant connectivity change due to coherence modulation, but we dismissed this due to the large variation in the direction of the effect, across participants. Considering the discussion above, the difference in connectivity between participants could be explained by differing neural responses of the regions involved. The conclusion of the above hypotheses would be that the apparent connectivity could increase or decrease as coherence varied depending on each region's response to different dimensions of the stimulus. Such inter-subject variability, if true, is intriguing, although this author feels it is more likely that the variability was probably due to unidentified problems with the DCM methods used or differences in ROI selection between participants.

Is there an auditory motion area?

As mentioned in the introduction, some studies have cast doubt as to whether a separate system representing auditory motion exists in the human brain (Smith et al., 2007). Although possibly confounded by the ROI selection, the decoding of motion in A1 suggests that A1 itself has motion-related activity, distinguished from merely being a component of spatial processing. The stimuli were carefully created to minimise confounds such as simple volume cues to ensure that any decoding was movement related and not to do with simple responses to IIDs.

The adaptation analysis, if it had been successful, was intended to detect auditory motion specific

processing. Our results did not find such adaptation in any modality, leaving the question unresolved. The earlier chapter's results, looking at the integration of visual and auditory self-motion cues found that they were integrated accurately, a task one might imagine would be quite challenging without specific auditory motion representation. Similarly, the presence of behavioural adaptation in the auditory-only condition may indicate neurons responding specifically to auditory motion. Although not induced in this experiment, it is widely reported that vection can be induced by auditory cues alone. It may be that the movement of auditory cues is indirectly processed by a non-motion-related cortical region which then feeds-back to self-motion regions. Such a tortuous route may explain the latency usually reported before vection is induced and the variability between individuals.

This study is relatively unique for its complex auditory stimuli, and the attempts made to avoid simple features of the stimulus from confounding the result. For example, the auditory cues were placed and moved in such a way that the intensity of the cues would remain roughly constant at the start and end of the trial. The sense of motion was achieved through the combination of changes in spectrum filtering, ITD and IID for each cue, across several cues. This is in contrast to studies such as (Smith et al., 2007) which use single auditory motion cues. Such single-use cues make such studies vulnerable to the claim that the classification or BOLD responses are mainly due to the confounding effect of intensity. Alink et al. (2011) approached this problem during analysis, by subtracting the mean of each clique's timecourse to ensure the univariate response in one hemisphere would not affect classification. It is possible in Smith et al. (2004) that although the response of the region was not increased during motion (compared to stationary auditory cues), there could still be auditory motion processing, only discernible through classification, or similar. Finally, even the processing in Alink et al. (2011) may not be sufficient, as one might hypothesise the variance of the signal to increase with the univariate response, a component uncorrected by the subtraction of the mean. Research in this field will need to develop auditory motion stimuli as confound-free as possible. It is hoped that this study offers an example of such a stimulus set.

Contrast and Coherence

The switch to high coherence partway through the experiment had less of an effect on decoding, adaptation or DCM than expected. The low visual contrast, brief presentation and low speeds may all be contributing factors. I would suggest that much more coherent stimuli are used in future. An alternative type of self-motion with much more extreme stimuli are rotations. This would also allow a continuous, block-like, design to be used, with 20 seconds worth of rotation in one direction or the other. However, the sense of rotation may be reduced by the participant's supine position within the scanner, also the use of forward moving stimulus was chosen in response to the findings of heading-sensitive neurons in MST and VIP in electrophysiological research, and an attempt to keep the stimulus more realistic. However, Duffy & Wurtz (1991) found that planar motion may be the stimulus which elicit the greatest response in MST, a result we possibly should have considered more carefully, as it may have applied to VIP.

Our original choice of low coherence was driven partly so that we could test for connectivity changes

using DCM or SEM. It was also based on the results of the psychophysics study and finally by research which found an increase in the BOLD response to low coherence stimuli (compared to high coherence stimuli). For example Giaschi et al. (2007) found that low coherence dot motion produced a greater BOLD response in hMT+. fMRI appears to be more associated with local field potentials (LFP) than with action potentials. An electrophysics study by Nauhaus et al. (2008) found that lower contrast visual stimuli induced the greatest LFP (possibly due to modulation of lateral connectivity), therefore it seemed reasonable to suspect that the low coherence stimuli would have produced the greatest BOLD response.

Attention

It is intriguing that it was the low-coherence group that had the greatest decoding scores in the visual areas. With such a small sample size it is dangerous to draw firm conclusions, but such an effect has been reported previously (Giaschi et al., 2007). It may be due to the increase in BOLD response for low coherence stimuli, mentioned above. However an interesting alternative explanation might be that the effect is due to attention. The low-coherence stimulus may require considerably greater attention to perceive. Previous studies (such as Jehee et al., 2011) have shown that additional attention can increase the BOLD response to attended-features. Thus this could explain the increase in decoding scores. To further study this, one would require a distractor task or dimension to direct attention towards or away from the feature of interest, mirroring the earlier results of Kamitani & Tong (2005).

Is VIP an integration area?

The experiment was designed to answer this question in three ways. First, adaptation was intended to test whether individual neurons within VIP respond to stimuli of either modality. This would have been a clear indication of the region's multimodal activity. Unfortunately unimodal adaptation was not found, which means the absence of cross-modal adaptation cannot be used as evidence against the region's place in integrating stimuli. Second, connectivity between VIP and two unimodal sensory regions was expected to be modulated by coherence. As mentioned previously, the RFX DCM results are equivocal, suggesting possible modulation, although not in the way we expected. Thirdly, and most tenuously, the response of the region to the two unimodal conditions was considered. VIP appeared to strongly respond to the auditory and bimodal conditions, but its response to the visual stimulus was considerably weaker.

Advantages of classification over adaptation

Adaptation analysis has the potential to provide strong evidence of integration; cells responding to one stimulus can be shown to also respond to a second. Cross-modal classification results, if they had been successful, would only provide indirect evidence of integration. However, an intriguing potential of the classification results is deciding whether to classify on the actual stimulus or the perceived stimulus (determined from the direction chosen by the participant). If classification had been successful the study would have allowed the perceptual and actual classification signals to differ. This had the potential then

to allow an investigation into which regions these differ, in which would shed light on where integration occurs and potentially how signals of different weights are combined.

Some studies already look at the difference between the stimulus and perception during classification; not all the information represented in the brain is necessarily available to act upon (Williams et al., 2007). A final potential feature of these types of analysis is the possibility of inter-subject differences becoming more apparent. This would be particularly interesting with regards to older/younger subjects or those with cognitive impairments.

Two final caveats are obvious, although easily forgotten. First, only some features of the neural response are decoded during such fMRI studies, and as the study cited above notes, not all neural responses are used behaviourally. Second, because the organisation of regions varies, comparing classification between regions is very difficult (for example the strength of classification does not necessarily depend on a region's involvement in the processing of that stimulus dimension). A final aspect which may change during fMRI experiments is a transition from encoding to decoding memories. The associated BOLD responses to which may differ (Naselaris et al., 2011).

Relation to the Psychophysics Experiment

The two key predictions offered by the psychophysics experiment are related to the changes in connectivity and the level of adaptation in different regions.

Connectivity Several models of integration were mentioned in the previous chapter, from the purely descriptive, for example, Bayesian and probability summation to neural, such as the Normalisation model and the model by Ma et al. (2006). The descriptive models don't offer any predictions regarding the underlying architecture performing the calculations. The Normalisation model and the Ma et al. model do. The key difference between these two models is that the weight of the connections between the unimodal regions and the integration region will vary in the Normalisation model. The previous chapter did not test whether the psychophysics results could be explained by the Ma model, although they presumably could be. The initial DCM analysis suggests the weights do change - supporting the Normalisation model. However, the direction of connectivity change varied between subjects suggesting this result was not robust. SEM also found no significant difference. It is difficult to argue that the Normalisation model is by necessity incorrect, as the absence of a robust change in connectivity may simply be due to problems with the design or stimulus.

Adaptation The region VIP appears to experience adaptation to stimuli of either modality and cross-modally (so a trial of a visual stimulus following an auditory trial has a lower BOLD response in VIP than the same trial following a null condition). Differences in adaptation between ROIs depends greatly on the size and accuracy of the ROI defined by the localiser, hence no statistical tests have been used to compare the results between regions. In simple terms the psychophysics results predicted that the bimodal region would have a lower adaptation effect than the unimodal regions, although this might be impossible to detect as the adaptation detected in the integration region (VIP) may be inherited from the

earlier unimodal regions. In general it seems there is more adaptation in the unimodal areas, so it seems that the model does have some support. It is unclear how significant this difference is though.

Superadditivity A final difference between the Normalisation model and Ma's is whether superadditivity would occur. There was no evidence of it being present, although the confidence intervals of the unimodal and bimodal responses in VIP could accommodate superadditivity if it exists. Therefore we can not make any firm statement about which model is more likely on the basis of sub- or superadditivity.

4.5.3 Future Experiments

Stimuli

The experiment found that the stimuli used did not generate sufficiently distinguishable neural activity to decode the direction of motion of either the visual or auditory stimuli (beyond a very minimal success in V1). This failure meant that many other aspects of the study were compromised. Similar problems were caused by the very low BOLD response, particularly to the adaptation analysis. In hindsight a different stimulus should have been used. Increasing the contrast, speed and coherence of the dots would be the three key changes to make to the visual stimuli. For the auditory stimuli I would recommend switching the whole study to a rotating stimulus, which will be easier to generate and perceive, with the additional benefit of being possible to use in a long trial. A compromise between the linear and rotating stimuli might be a trajectory on a circular path, such as used by Diekmann et al. (2009).

Adaptation

A repeat of the adaptation experiment conducted here, but with more successful stimuli, would be very interesting to investigate VIP's importance as a region for integration. I would suggest returning to the more reliable, if less naturalistic stimuli, used in earlier auditory and visual motion experiments. For example, rather than rely on a simulation of forward motion, the simulation of rotation may be easier. This would mean the visual stimuli would be simple translating dot-motion (for example), while the auditory stimuli will be a continuously rotating set of cues rotating 'around' the person. Although not stated explicitly, studies using a rotating auditory stimulus appear to be more successful at inducing vection compared to linear vection studies, e.g. comparing the results of circular (Larsson et al., 2004) and linear (Sakamoto et al., 2004) studies, and the associated visual stimuli are easier to simulate and more plausible, perceptually (anecdotal, with participants during experiment development) Such plausibility may be important, as it has been found that vection is more successful if participants believe there is a potential for actual motion (Riecke et al., 2009). A rotating stimulus would also be more apt for using in a blocked-design. Not all 128 trials would fit within one session, but it would be possible to perform the adaptation experiment by splitting them into several sessions and repeating the last few trials of the previous session in the following session. The longer trials will both increase adaptation if it exists, and

increase the power of the experiment (by increasing the signal-to-noise ratio). The voxel size should be increased considerably (to 3mm or 4mm), to further increase the SNR. Finally, 100% coherent stimuli should be used. The participants should still be required to do a task associated with the stimulus, for example some form of speed judgement would be appropriate. These changes will give the future experiment considerably more power in detecting adaptation.

Adaptation and Classification

An alternative way to investigate adaptation would be to look for more subtle changes in the activity of the region by using a pattern classification method. A classifier would then be required to compare those trials following the same direction over those trials following the opposing direction. Any detected differences would (assuming the experiment was properly counterbalanced) be due to the adaptation of the region. Classification methods appear to be considerably more powerful, and could be more robust to poorly localised ROIs.

4.5.4 Conclusion

Although some intriguing hints about the processing of self-motion cues have been reported, overall the experiment did not achieve the majority of its aims. One reason is that the design of the experiment was split between several objectives, with conflicting demands on the experiment design. For example the low coherence stimuli were required for the connectivity analysis, but may have reduced the signal when testing classification and adaptation. Similarly, smaller voxels seem wise for classification but will have reduced the signal-to-noise ratio for the adaptation and connectivity analyses. The choice of relatively brief trials was driven by a perceived need to maximise the number of trials for the adaptation analysis, but may have weakened the individual trial's signal unacceptably. Finally, the use of a circular self-motion cue, although less novel than linear motion, could have offered a considerably greater sense of motion and simplification of the study. It is hoped future experiments will be performed which build on these results and suggestions.

Chapter 5

Head Direction

Some of the background to this chapter is reproduced in a shortened form from my MSc research project (Smith, 2009), recently incorporated into Bett et al. (2013).

The experiments described in previous chapters have all used simple stimuli that specifically exclude landmarks and other such navigation features. These experiments were intended to investigate some of the earlier stages in the self-motion processing pathway. For example the dot motion stimuli emulate optic flow which is particularly salient for inferring self-motion. An important and related aspect of navigation is the constant updating of our orientation within our environment through the use of self-motion cues. This chapter investigates these more abstract layers of spatial processing, and looks at the representation of both place and orientation in several cortical areas.

The most likely candidates for regions responsible for tracking our orientation and location lie in or near the hippocampal formation, a belief based largely on the discovery of place, grid and head direction cells, recorded from behaving rats. The research into such cells has been largely constrained to non-human animal models. If one could reliably decode or detect the activity of these cells in humans, using fMRI, then considerable progress could be made in understanding their responses to both intention, attention and their interaction with perception, without the need for invasive surgical procedures. All questions difficult to investigate in animal models.

This experiment's hypotheses were limited to attempting to decode the direction and place of the participant (in a virtual environment) from regions of the brain believed to contain relevant cells. The aim was to first show such decoding was possible and second to develop and improve the methodologies involved in such analyses.

5.1 Background

5.1.1 Connections and Overview

Below I've tried to summarise the most salient and relevant aspects of research into the navigation system. This includes the hippocampus, many neighbouring regions and various thalamic nuclei. Figure 5.1 is an overview of the connectivity in the navigation and spatial processing circuits of these structures.

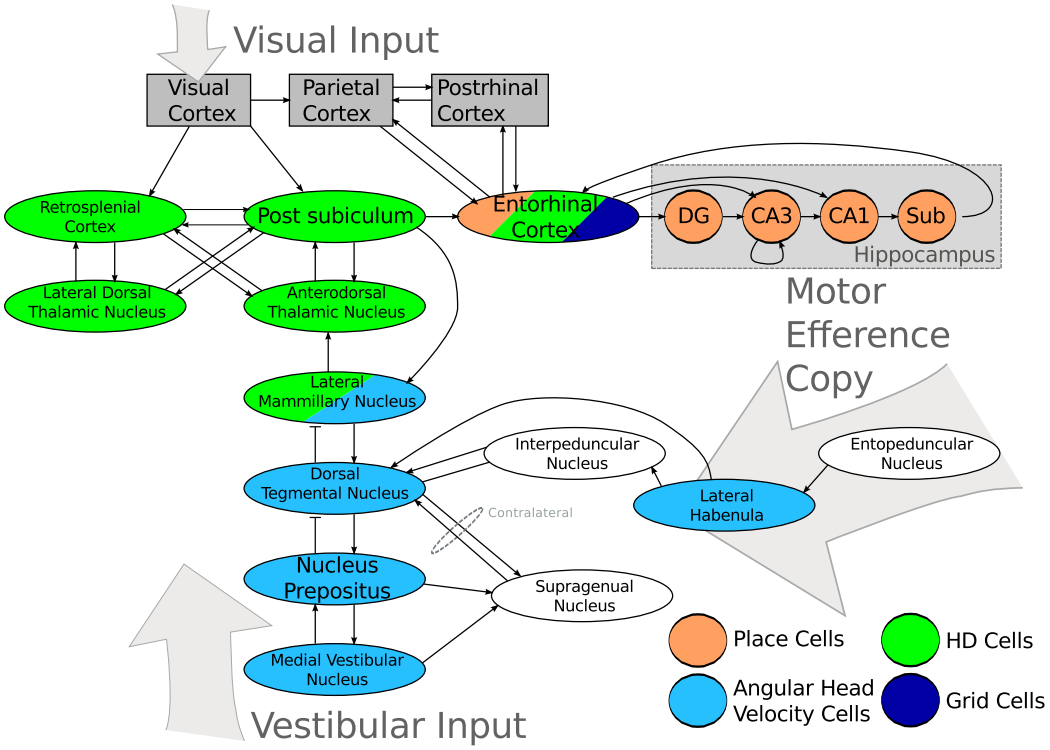


Figure 5.1: Hippocampal Formation and associated structures. The figure shows the likely connections between fields of the hippocampal formation and other relevant areas. Data from Goodridge & Taube (1997, fig. 1), Sharp et al. (2001, fig. 4), Andersen et al. (2007, p38, p108) and discussions with Emma Wood. Abbreviations: DG, Dentate Gyrus; CA3 and CA1, subfields of the Hippocampus; Sub, Subiculum. For Retrosplenial cortex connections see Vann et al. (2009).

In general there are three inputs into the navigation processing brain regions. These are from the vestibular sense, motor efference copy and vision. Other sources (such as auditory cues and the tactile sensation of objects and environmental features) appear to play a part too. Associations between spatial orientations and locations are integrated and associated with particular head or place cell activity through some form of local plasticity. My master's project (contributing to Bett et al., 2013) for example, investigated the importance of the post-subiculum in mediating such plasticity for the head direction system.

5.1.2 Head Direction

Head Direction Cells

Head direction cells are defined by their increased firing rate when an animal's head is orientated in a particular direction. They were first discovered in the post-subiculum (dorsal presubiculum) (Taube et al., 1990), but have since been found in several other areas (Andersen et al., 2007, p521) including the retrosplenial cortex, some thamalic nuclei and the entorhinal cortex (see Figure 5.1). Paths of information through these systems were determined by lesion studies and cell recording studies. For example, head direction cells in the post-subiculum were found to still function after bilateral hippocampal lesions (Golob & Taube, 1997). However, when the postsubiculum were lesioned, place fields were lost and did not return (Calton et al., 2003).

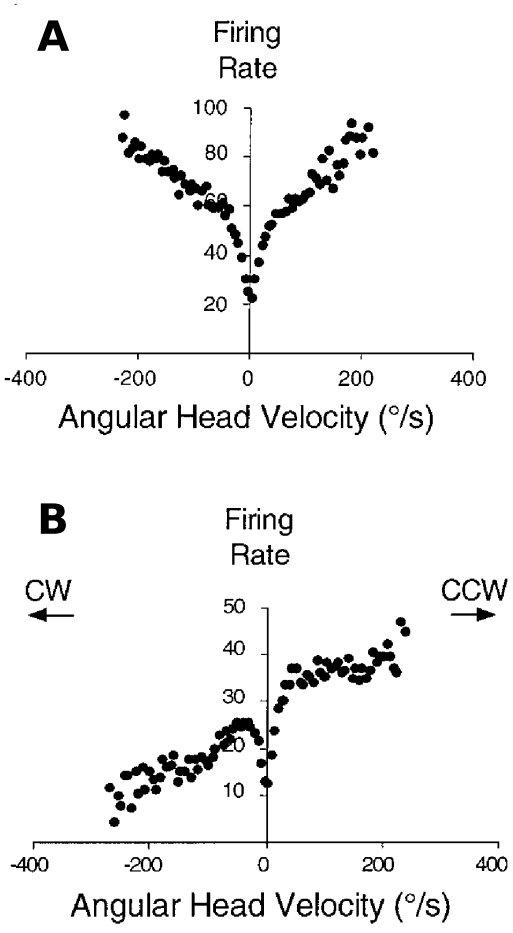


Figure 5.2: Examples of Angular Head Velocity Cells, reproduced from Bassett & Taube (2001). A. Symmetrical Angular Head Velocity Cell. B. Asymmetric Angular Head Velocity Cell.

Taube et al. (1996) reported that head direction cells have limited adaptation. This means the BOLD signal from the region may not be overly influenced by repeated presentation of the same stimulus direction, and so an adaptation design might not be successful. Hence the study's choice of a pattern classification design.

Vestibular Input

Bassett & Taube (2001) found that cells in the Dorsal Tegmental Nucleus of Gudden (DTN) fire at different rates depending on the speed the rat rotates its head. It seems that the vestibular input is represented at the lowest levels by the activity these cells, dubbed angular head velocity cells.

The integral (over time) of the output of the angular head velocity cells will give the head direction of the rat. This process appears to occur in the Lateral Mammillary Nucleus (LMam). Stackman & Taube (1998) found head direction cells in this region, which appeared to fulfil this integrating roll. It was supposed that a form of continuous attractor, possibly in the Postsubiculum (Stringer et al., 2002), performs this integration.

Importantly, these regions, as far down the hierarchy as LMam, have reciprocal and feedback connections (Goodridge & Taube, 1997). It is these connections which explain how a visual stimulus can influence such head direction cells.

Motor Efference Copy

Motor Efference Copy is probably incorporated at the level of the dorsal tegmental nucleus, although other paths exist for this information to reach the hippocampus (Taube et al., 1996). In our experiment such an input is unavailable; fMRI studies require participants to remain stationary throughout the scan. This, at first, seems incompatible with the study of spatial processing and the analysis of head direction signals. We must ask whether the relevant cells will continue to fire as expected in a such a situation? Several lines of enquiry suggest that the cells will respond correctly to the visual environment cues alone. First, the activity of head direction cells in rats while they are passively rotated allows us to see what effect removing motor efference copy has on the cells. Taube (1995) found that the effect of passive rotation depended on which region the head direction cell recordings were from. In the ATN, HD cells appeared to be very attenuated, while in the postsubiculum, the cells appeared to fire in a similar way to the non-passive case. Recent fMRI studies in humans found evidence of both grid (Doeller et al., 2010) and place-cell (Hassabis et al., 2009) related activity used completely passive viewing. This suggests that motor-efference copy is not necessary for those cell types. It is believed, based on evidence from lesion studies (Calton et al., 2003), that head direction cells in the post subiculum and elsewhere are essential for the functioning of grid and place cells in rats. It therefore seems likely that most of the populations of head direction cells (besides those in the ATN) will respond well to a purely passive-viewing stimulus.

Visual inputs are known to play an important part in the responses of head direction cells. For example, rotating a cue card around an enclosed cylindrical environment causes the head direction cells

to rotate their preferred direction. In other words, the head direction cell responses will generally rotate with the most salient cues (Knierim et al., 1995), with visual stimuli overriding others (Goodridge et al., 1998), even when these conflict with all idiothetic cues. Place and grid cells rotate in a similar way [1]. This would suggest our passive viewing design will influence head direction cells, even though the participant will not be experiencing the associated vestibular or motor-efferent cues.

As will be described in the methods section, participants are only exposed to the environment for 20 or so minutes prior to the fMRI scanning. Therefore one must give consideration to the question of whether head direction cells would be associated with the environment. Goodridge et al. (1998) noted that 'only 8 min of exposure to a salient visual cue is enough time to enable it to exert stimulus control over HD cells'. This fast plasticity means the virtual environment stimuli should be able to associate head direction cells with environmental directions within the training period.

A final relevant study is the impressive patch clamp recordings of place cells made from a mouse while it ran on a spherical treadmill, while viewing a toroidal screen which covers almost its whole visual field (Harvey et al., 2009). All but the vestibular cues to self-motion were available to the mouse, and the place cells functioned as expected. The artificial nature of the visual scene did not appear to perturb the responses, suggesting artificial, computer generated environments do influence place cells in a similar way to real environments.

From the discussion above, it is clear that the areas the study should focus on for detecting a head direction signal are the retrosplenial cortex and several portions of the thalamus (e.g. the Anterodorsal Thalamic Nucleus and the Lateral Dorsal Thalamic Nucleus). The other area likely to be heavily involved in the head direction system is the post subiculum (see Figure 5.1). However, this region is very small, making a decoding fMRI study very difficult, as only a very few voxels will include the region and the accuracy of localisation will be in doubt.

5.1.3 Place Cells and the Hippocampus

Hippocampus Background

The hippocampus is an elongated structure which, in humans, extends along the medial-ventral surface of the temporal lobe (Figure 5.3). Historically many functions were assigned to the region, including olfaction, motor function, reason, stubbornness and inventiveness (Andersen et al., 2007, p9). It wasn't until the pioneering surgery performed by William Scoville on patient HM (Henry Molaison) and Brenda Milner's detailed case studies which followed, that we learned something of its purpose. HM's amnesia provided the first and most shocking evidence of the importance of the hippocampus for memory formation. It is hard to imagine now that such a link was unexpected, but in the words of Scoville and Milner, 'There has been one striking and totally unexpected behavioural result: a grave loss of recent memory in those cases in which the medial temporal lobe resection was so extensive as to involve the major portion of the hippocampal complex bilaterally.' (Scoville & Milner, 2000).

[1] Noted in literature and from my own observations during my MSc project, see also Bett et al. (2013).

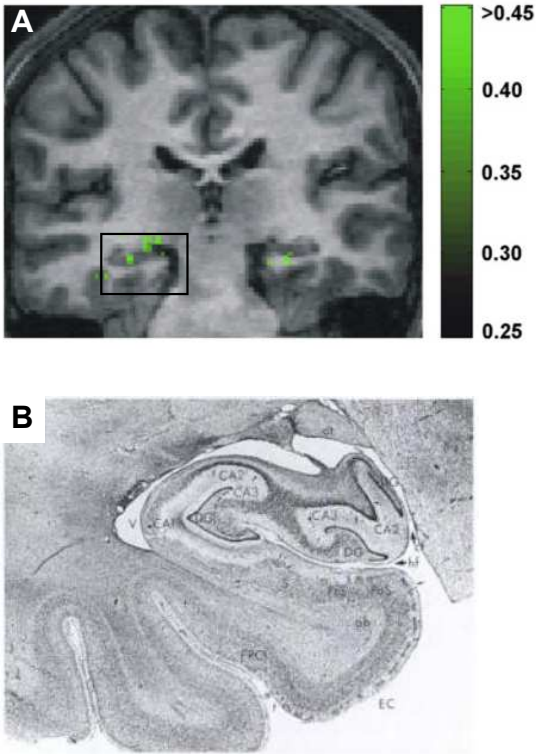


Figure 5.3: The location of the human hippocampus. A. fMRI T1 anatomical image highlighting regions in which place decoding was significantly above chance (Hassabis et al., 2009). B. Nissl stained coronal sections through the human hippocampus (Andersen et al., 2007, p105).

Place Cells

It was the introduction of cell recordings in a behaving animal which led to the more recent findings surrounding the hippocampus (Figure 5.4) and its relevance for spatial processing and memory. Possibly the most famous result was by O'Keefe & Dostrovsky (1971) who found individual cells responding to the location of the animal. Dubbed place cells, they were originally found in CA1 (a subfield of the hippocampus), and generally fire independently of direction. Although not proving causality, the coherence of place fields has been found to be correlated with performance in map-based navigation tasks (Lenck-Santini et al., 2002). Lesion experiments using the Morris water maze found spatial memory particularly affected by lesions of the hippocampus and many of the surrounding areas (Miller & Vogt, 1984). These findings fit well with the discoveries of spatially responsive cells.

In the study by Ekstrom et al. (2003), human place cells were recorded directly. Prior to surgery for epilepsy, several patients had cells recorded while playing a taxi-driver computer game, in which they explored a virtual town. The study found that humans, too, have place cells (mainly in the hippocampus) and other cells which respond to views of landmarks (in the parahippocampal region).

The current study was based in particular on the findings of an fMRI study (Hassabis et al., 2009)

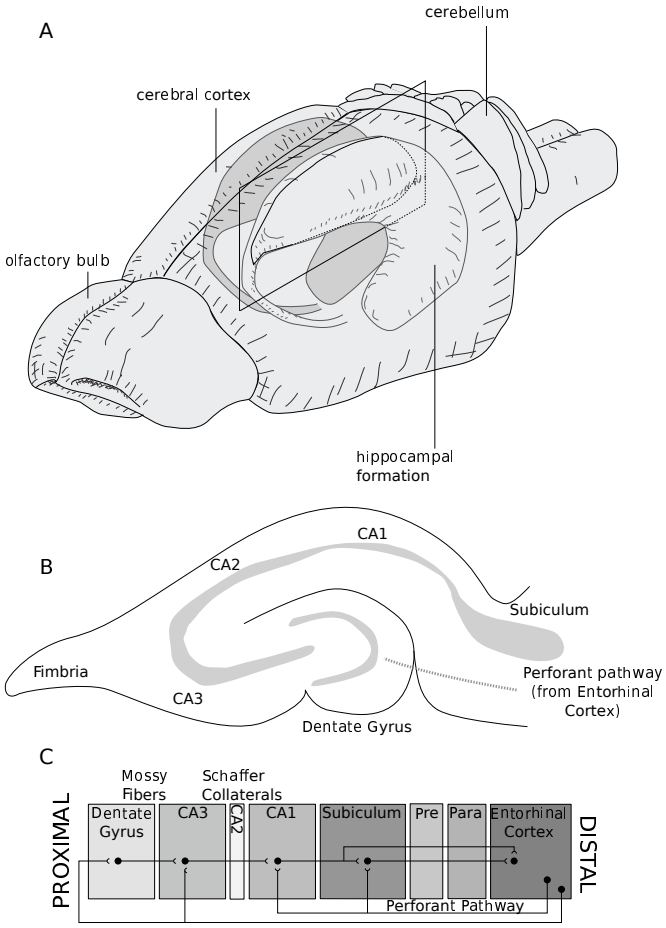


Figure 5.4: The hippocampal formation in rats: A, Location of the Hippocampus, based on Cheung & Cardinal (2005, p12, fig3) and Andersen et al. (2007, chp3). B, Saggital section through the hippocampal formation, from Andersen et al. (2007, fig3.14). C, Connections within the hippocampal formation in a plane outwards.

which decoded the position of a participant in a virtual environment using the ensemble activity of multiple voxels recorded from the hippocampus. Participants were required to navigate between four locations in a square room. A limited number of landmarks allowed the participants to orient themselves within the environment. Several measures were taken to avoid confounds caused by the visual stimuli. In particular the view was rotated to face downwards each time the participant reached a target. A searchlight analysis was performed on the fMRI data to find those neighbourhoods of voxels whose ensemble activity encoded the location of the participant (between either two or all-four of the targets). Multiple comparisons were made (due to the multiple searchlight locations, one for each of the 6750 voxels within the bounding-box). Rather than correct directly for multiple comparisons, the authors compared the proportion of spheres that have significant results, between the two regions. They found the hippocampus has more significant results than the parahippocampal region when decoding place,

and the converse when decoding which room the participant was exploring.

Since their original study, a second study has decoded place-information from the hippocampus. In Rodriguez (2010), participants were required to navigate to a previously visited location in a virtual Morris water maze task. In this austere environment, only one orientation landmark was provided (a red square was drawn on the wall). Intriguingly, the hippocampus was found to provide the most salient and unique location information of the regions decoded. Rather than using a search-light decoding method, Rodriguez (2010) used a partial least square classifier. Of the eleven participants, four were rejected for not achieving a performance threshold of 7 out of 12 correct for the North goal. Of these seven, all achieved above chance performance.

Both searchlight and standard decoding analyses were performed, as it was not clear which would be most successful.

A more recent fMRI study looking at the ensemble response of the hippocampus and parahippocampus to place cues, used still images from highly-familiar real-world locations (Morgan et al., 2011). The study found adaptation effects for nearby landmarks in the left anterior portion of the hippocampus. An important result was the decoding of landmark identity from several regions. The PPA, RSC, LOC and early visual areas, all successfully decoded the landmark identity above chance. However, the four hippocampal regions did not provided detectably significant information about landmark identity. This absence of place-specific information compared to the previous studies does raise questions regarding what signals the previous studies were recording. It might be that the absence of a goal-related location meant the place-associated activity was absent. Alternatively the absence of actual navigation meant the place-activity reported in the other studies was not available. In their conclusion Morgan et al. (2011) suggest the responses may be due to episodic memory encoding and retrieval, rather than the place-cell related spatial navigation of the other studies. Supporting this conclusion is an older fMRI study by Strange et al. (1999), looking at adaptation and the response to novelty in an item-learning task. In this study the left anterior hippocampus adapted to both perceptual novelty and semantic novelty. The general nature of the region's adaptation suggests the results in Morgan et al. (2011) may not be particularly spatial and may instead be related to the strength of association between the landmarks used.

An earlier, related study (Spiers & Maguire, 2007) found that the distance to a goal location were correlated with the BOLD signal (positively) in the medial prefrontal cortex and (negatively) in the right subiculum (part of the hippocampal formation), and the right entorhinal cortex. The orientation of the goal location was found to be correlated with activity in the posterior parietal cortex. The correlation of distance with the BOLD responses seems to mirror the results in Morgan et al. (2011). Whether these correlations are related to the decoding described in the other papers is more difficult to determine.

Grid Cells

For completeness, it is worth mentioning grid cells. First discovered by Hafting et al. (2005) in the Entorhinal cortex, they fire only when the rat is at a vertex of a triangular grid, the scale of which varies across cells. An fMRI study conducted by Doeller et al. (2010) made clever use of the fact that the

grids of grid cells are aligned. Motion along the grid's edges will involve passing over vertices more frequently than motion 60° offset. Due to grid alignment, macroscopic effect becomes apparent, with the average entorhinal cortex fMRI signal being greater when in one of the six directions aligned with the grid compared to the six directions misaligned. The study's use of visual-only stimuli provides additional support for our own visual-only design.

5.1.4 Retrosplenial Cortex

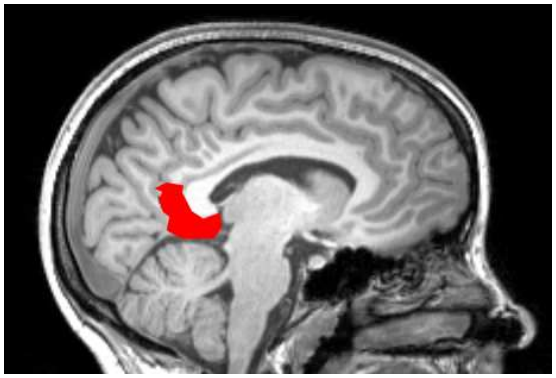


Figure 5.5: The approximate location of the Retrosplenial Cortex.

From the name, one knows the the retrosplenial cortex lies posterior to the splenium (the most caudal part of the corpus callosum). The retrosplenial cortex is part of the posterior cingulate cortex. It comprises areas 29 and 30. Area 29 is a region within the callosal sulcus (the sulcus between the corpus callosum and the cingulate gyrus). Area 30 extends from the callosal sulcus to the bank of the cingulate gyrus.

Animal lesion studies appear to suggest the retrosplenial cortex has an important role in spatial memory (review Vann et al., 2009). In particular lesions of the dysgranular region, which is known to have more interconnections with visual regions, caused impairments in the use of visual cues. More recently, electrophysiological recordings in rats, directly from the region, have found 8.5% of the cells recorded were head-direction cells (Chen et al., 1994), emphasising its relevance to navigation and spatial processing. The review by Vann et al. (2009) suggests that the retrosplenial cortex is a key pathway for visual information to enter the head direction (and spatial processing) system. It also collates findings from lesion cases which appear to show the importance of the retrosplenial cortex in human memory and orientation. An earlier fMRI study by Epstein et al. (2007) puts emphasis on the region's importance in spatial navigation. In particular they hypothesise that, while the Parahippocampal place area responds to spatial structure, the retrosplenial cortex is more associated with place recognition. The active-navigation study by Spiers & Maguire (2007) reported increases in activity in the retrosplenial cortex when spatial processing or learning were underway.

Besides these intriguing spatial-associations with the region, for this experiment it is the presence

of head direction cells in rats which makes the retrosplenial cortex a candidate region of interest for decoding head direction.

5.1.5 Other regions

FFA

If the location or direction of the participant is successfully decoded from the fMRI signal in one of the above regions, it might be argued this is due to a non-spatial stimulus. To control for such hypothetical confounds the Fusiform Face Area (FFA) of the brain will be used, as we expect no spatial processing there. Instead, the FFA has a greater response to faces than to other stimuli Kanwisher et al. (1997), although there is still debate about whether the area is truly specific for faces (e.g. Hanson et al., 2004).

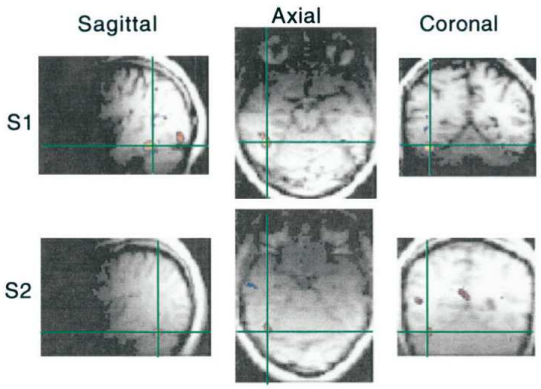


Figure 5.6: The localisation of the Fusiform Face Area, from Kanwisher et al. (1997).

Parahippocampal Place Area

As mentioned in section 5.1.3, direct recording from cells in the human parahippocampal region found cells which respond to particular views of landmarks. Several fMRI and behavioural studies followed this up. For example Epstein et al. (2007) reported that the Parahippocampal Place Area responded equally strongly to all the four types of environmental scene used, but responded more weakly when the stimulus was the picture of an object.

Epstein et al. (2007) summarise that, for the PPA, what is most important is that a scene is viewed, not what type of information is required or task is being performed. The fMRI BOLD contrast between presenting images of rooms and faces allowed us to replicate the original PPA localiser from Epstein & Kanwisher (1998) (Figure 5.7).

This experiment's place and head direction decoding tasks were carefully developed to avoid any confound caused by the same visual object appearing reliably in one direction or location. Ekstrom et al. (2003) found 'view-dependent' cells (mainly in the PPA). These cells fired preferentially when the

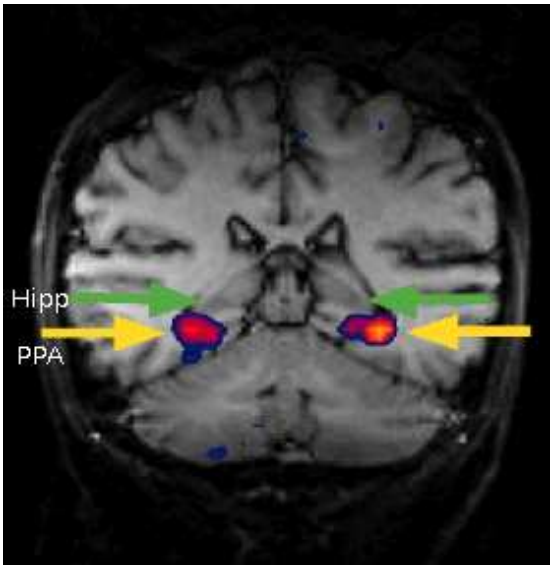


Figure 5.7: The localisation of the Parahippocampal Place Area, from Epstein & Kanwisher (1998), using a similar room image stimulus as our study's localiser.

participant was facing particular shops. It would seem likely therefore that our experiment should avoid generating a reliable signal in the PPA for particular locations or directions.

Visual Areas

The middle temporal visual complex (hMT+ or V5) was described in detail in Chapter 2. The region is known to respond to visual motion, a likely precursor in the generation of rotation information for the head-direction system. In this experiment however, the region was selected for analysis as a separate control to confirm the decoding, stimulus and processing were all functioning as expected. A failure to decode direction of motion could indicate the analysis was at fault.

5.1.6 Summary

With studies finding evidence of both grid cells and place cells in humans, it seems reasonable that head direction cells could be found in a similar way.

In the previous studies a visual-only stimulus appeared to be sufficient for the cortical regions to respond as if real motion and spatial orientation was taking place. An important aspect which this study will try to duplicate is the importance of self-driven motion, rather than passive viewing.

The main aim of this study was to decode head direction information from those locations in which we expect to find head-direction cells (based on animal studies) in humans. Simultaneously, the experiment should be expected to reproduce the place-decoding results of Hassabis et al. (2009).

In summary, for place activity, the obvious regions to investigate include several subfields of the hip-

pocampus. As fMRI has limited spatial resolution, the whole region will be included as a single region of interest (ROI). Morgan et al. (2011) divided up the hippocampus into subregions, and found only one portion (the left-anterior hippocampus) to have adaptation. Our whole-region decoding used LARS classification which is particularly useful when classifying a high dimensional input. Our analyses also uses a searchlight procedure, so that if a subfield of the region was particularly important, it would still be found.

5.2 Methods

5.2.1 Pilot study

Earlier versions of the environment and task were tested before settling on the current design. In the initial environment the landmarks were more numerous, complex and non-symmetric (Figure 5.8). The experimental design was presented at the CCNi where useful feedback was given. The main concern expressed related to the visual confound caused by the cues which would be seen in the same direction from two or more of the target locations. Even though the 'mist' will hide the cues during the key volumes, various mechanisms may exist to allow 'visual storage'(Parks, 1965). A recent study (Smith & Muckli, 2010) found that parts of V1 and V2 with inputs from occluded sections of the visual field contained decodable information about the visual stimulus presented to the rest of the visual field. To avoid such confounds the distal cues were placed so that the participant would be unable to see them when facing in any of the cardinal directions. The local cues were also organised so that they would be generally not be visible in the sixteen conditions. Those that would be were placed so that they would be visible in a different direction from each of the target locations (and are visible in only 5 of the 16 conditions). These changes made the task considerably more difficult and extra training was found to be necessary.

Figure 5.8: An earlier environment tested during a pilot study (v4).

5.2.2 Participants

Ten participants took part in the study, aged 20-31 (mean 22.7, standard deviation 3.5, 5 female). All reported normal or corrected to normal vision. Participants were recruited from the University of Glas-

gow's Centre for Cognitive Neuroimaging (CCNi) subject pool database. The study was approved by the ethics committee of the CCNi.

5.2.3 Apparatus

fMRI Scanning

BOLD signals were recorded using the fMRI scanner at the university of Glasgow (Siemens 3T TIM Trio, using a 32-channel head coil). We acquired approximately 800 volumes in each of the 3 sessions (although this depended on the speed the participants completed the tasks) and 150 volumes in the localiser. Each volume had 30 slices, with slices of $2mm$ (gap $0.2mm$), with $2mm \times 2mm$ voxels. Figure 5.9 illustrates the volume scanned in a typical subject. It includes the entire temporal lobe, the hippocampus and surrounding hippocampal formation, the retrosplenial cortex and the visual cortex.

The anatomical scan was a standard T1-weighted MRI sequence (TR=1900ms, TE=2.52ms, FA=9°, $1mm^3$ isotropic voxels), taken during the same scanning session.

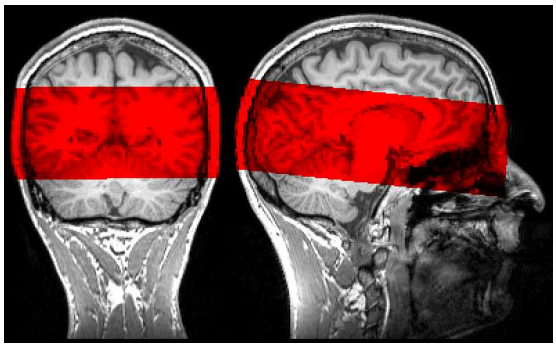


Figure 5.9: Coronal and Sagittal sections illustrating the region imaged during the experiment (subject ECN13).

Projector, Screen and Joystick

Visual stimuli were projected with an LCD projector onto a screen placed behind the subject. This was viewed using a mirror placed in front of the subject, angled at 45°. The visual stimulus was produced using Vizard 3.0 (Worldviz, www.worldviz.com). The screen appeared approximately 110cm from the participant's eyes, and extended approximately 18° horizontally and 14° vertically.

The joystick used was the fMRI compatible HHSC JOY 1 device (Current Designs, Philadelphia, PA, USA). The joystick also had buttons on the base, any of which were used by the participant to respond when necessary. At the start of the experiment a single message asked users to centre the joystick and press the joystick button. This ensured the joystick was properly calibrated.

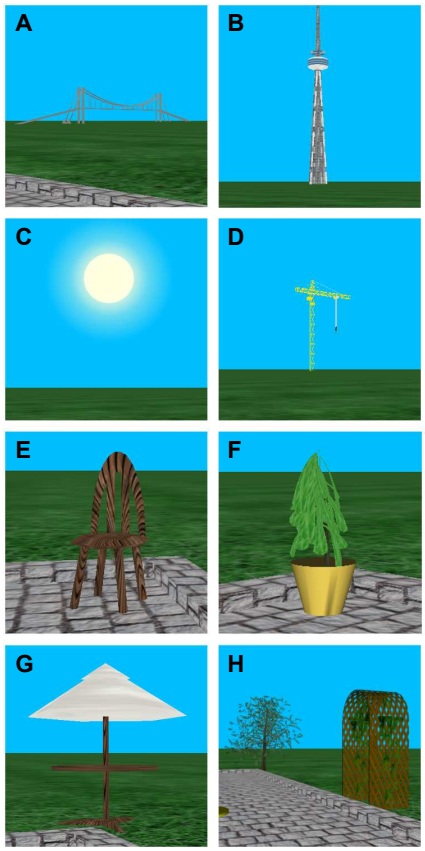


Figure 5.10: The landmarks from the fMRI experiment. The four distal landmarks; A. Bridge; B. Tower; C. Sun; D. Crane. The four local landmarks; E. Chair; F. Pot plant; G. Umbrella; H. Tree. The arch is also visible in the last image.

5.2.4 Stimuli

The visual stimulus projected on the screen consisted of a 3d virtual environment (see figure 5.12). Within the environment were four distant landmarks (rendered 'at infinity' to remove motion parallax cues, figure 5.10). The location of these four landmarks varied between subjects. In the example they were as follows; NW, a tower; NE, a crane; SE, a bridge; SW, the sun). The participants were restricted to move within a square patio area around a building. At each corner of the patio a local landmark was placed which did not move between participants (NW, an umbrella; NE, a tree; SE, a chair; SW, a pot plant, figure 5.10). An additional landmark (an arch) was placed to the East, between the chair and the tree to aid participants in keeping track of their location, as it was found learning of the environment was greatly improved with the addition of such a non-symmetry in pilot studies. In the centre of the square patio was a house (Figure 5.11) Surrounding the house were four circles drawn as raised yellow disks on the ground. These were placed in the four cardinal directions around the patio. Although unlabelled in the presentation, they were assigned labels the participant had to learn. This was intended

Figure 5.11: The house landmark from two angles, found in the middle of the patio.

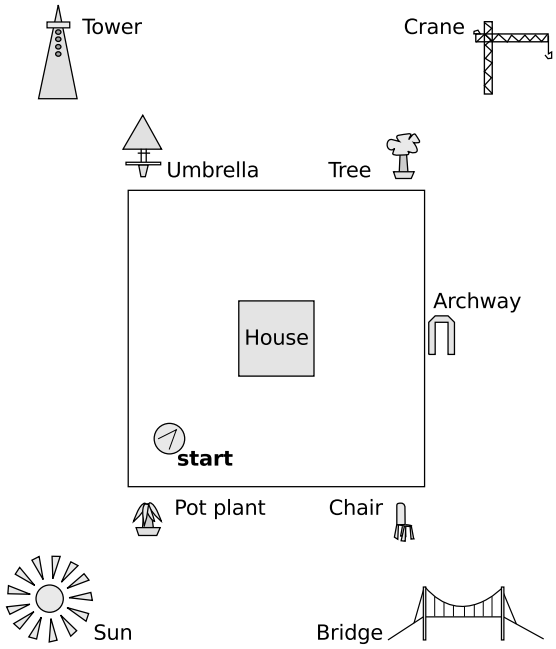


Figure 5.12: The whole environment. The four distal cues and the labels of the four targets could change between subjects.

to minimise the direct associations a participant might form with a physical label or object, which might confound any place-decoding findings, especially in a region of the brain such as the hippocampus, heavily associated with such associative spatial-memory. To further avoid this, besides the arch, all the landmarks were equidistant between targets.

The participants were able to move forward and rotate in the environment. They could move at a maximum of $3.75ms^{-1}$, which, although faster than walking speed is within a humans natural limits of speed (8.3mph, which equates to a moderately fast run). To reach this speed, the joystick had to be fully deflected forwards. Tilting the joystick to full-deflection to the left or right allowed the user to rotate in the virtual environment up to $25°s^{-1}$ (allowing a full rotation to occur in 14.4 seconds). The eye height within the environment was fixed at 1.8m.

5.2.5 Procedure

Instructions and Training

Prior to the experiment, the subject had 20-30 minutes training in which they learnt to navigate the environment, use a joystick and learn the locations of targets and landmarks. They were required, in each trial to travel to one of the four landmarks (A, B, C or D) as quickly as possible. The assignment of these landmark labels varied between subjects. Each trial began with the label of the target cue being displayed for 3 seconds. The target label remained in the corner of the screen throughout the trial until they reached the target. During training the participant was free to navigate around the environment, visiting each of the targets. If they visited the wrong one, a message appeared saying (for example), 'Wrong Location: D', where 'D' was the name of the location the participant had reached. In the experiment itself, this extra hint was removed, and the participant had to rely on memory to recall which location was which. If they arrived at the correct location, the message said 'Correct Location: C'. Again, 'C' was the name of the location, and was only included during training.

At this point, control of navigation was taken from the participant. Their position was adjusted slightly to the centre of the disk and then the direction they are facing was rotated at a rate of $40^\circ/s$, to face in one of the four cardinal directions. It was emphasised how important it was that they keep track of their orientation in the environment. These were chosen so that none of the distal landmarks were visible in the final facing direction. The intention was that the head direction would be made independent of landmarks. To allow the participant time to perceive their new orientation, a period of 1.6 seconds was permitted to elapse before a 'fog' descends over 1.2 seconds. Originally this was purely a grey plain fog-like reduction in visibility, with distant objects being masked first, followed by local object and finally the screen becomes a uniform light-grey. The view, and therefore any after-image, from each of the 16 location/direction combinations would be slightly different (even those facing out into the plane, due to variations in the texture of the grass), so an additional pink-noise texture was included whilst in the fog to mask any after-images.

In the training trials, after the mist descended, a random number of seconds elapsed (between zero and six), after which a message was displayed asking the user to 'Point to the X and push joystick button' (where X was one of the four distal landmarks). They were required to tilt the joystick in the direction of the requested cue, then press the button. After they had clicked the button a message is displayed (during training) saying either 'Correct' or 'Incorrect'. Then the fog was removed and the next trial started.

The training trials were divided into three blocks of 16 trials. The first was as described above. The second and third were similar but the participants were required to point the joystick and click within three seconds of the prompt. If they failed a message appeared saying 'Joystick Response Timed Out!'. During the experiment such feedback was not given.

Main Experiment

The main experiment was in the same environment (with the same landmarks, etc). The task was the same except for three changes. First, each block had 17 trials. Second, the participants only needed to do the pointing task in three of the 17 trials. Of these, one occurred earlier than six seconds. This was to ensure they were still 'thinking' about their orientation during the six seconds - something they might learn not to need to do if they were never tested. This probe trial was excluded from the classification algorithm. Only three trials were tested to save time during the fMRI scanning, so as many trials as possible could be included. The participants did not know which trials these would be, so were required to keep track of their orientation in every trial. The third change was that they were not given feedback on the identity of the four targets. By this point all the participants were expected to have learnt which one was which. Obviously only the correct target would trigger the mist and the end of the trial.

There were 8 initial trials in the fMRI scanner to allow the participant to familiarised themselves with the new screen, joystick and position. Then three scanning sessions were completed, each one lasting about twenty minutes. Within each there were four blocks, each completing the sixteen location/direction combinations.

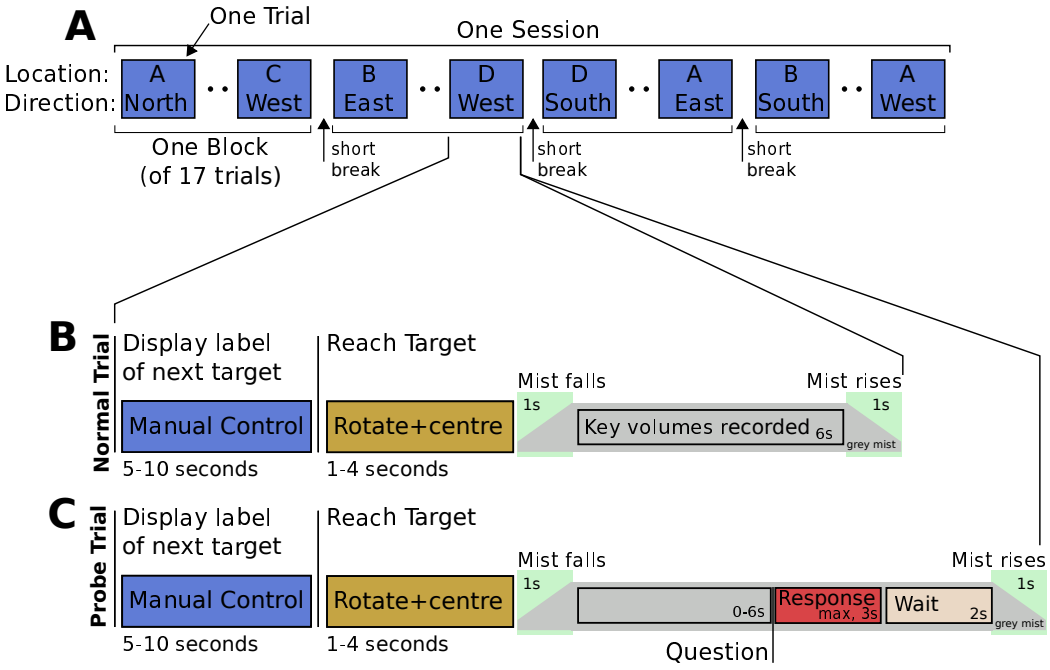


Figure 5.13: The experimental procedure. A. One session is divided into 4 blocks, each containing 17 trials. B. The normal trials consist of a period of manual control in which the participant navigates the environment, then a period of automatic movement in which they were rotated to face a particular direction. Finally the mist covers the view for 6 seconds; the period of most interest. C. The three 'probe trials' are the same but the participant is asked to point the joystick at one of the distal cues. On one trial this occurs prior to the end of the 6 seconds.

Region	Localisation Method	Expected Signals
Posterior Cingulate	Anatomical (freesurfer)	Head Direction
Retrosplenial Cortex	Functional Localiser ('PPA')	Head Direction
Thalamus	Anatomical (Normalised MNI brain atlas)	Head Direction
Hippocampus	Anatomical (Normalised MNI brain atlas)	Place
Fusiform Face Area	Function Localiser ('FFA')	None
Parahippocampus Place Area	Functional Localiser ('PPA')	None
Middle Temporal (V5) Complex	Anatomical (Normalised MNI brain atlas)	Motion Direction/Mist
V1	Anatomical (Freesurfer)	Motion Direction/Mist

Table 5.1: Regions localised

5.2.6 Localisers

Seven regions of interest were localised (either anatomically or functionally, Table 5.1). The retrosplenial cortex was located through a functional localiser, but was also considered to be part of the posterior cingulate, a region localised using an anatomical toolbox.

To localise the Fusiform Face Area (FFA), the Parahippocampal Place Area (PPA) and the Retrosplenial Cortex (RSP) a functional localiser was performed at the end of the experiment. The localiser was modelled on the experiments which originally identified the PPA (Epstein & Kanwisher, 1998) and FFA (Kanwisher et al., 1997). Epstein et al. (2007) used very similar localisers and also identified the RSP. The localiser took 280 seconds, and consisted of 8 blocks of 22 image presentations. Each image was displayed for 400ms, with 480ms periods between. The blocks were of 20 building interiors or 20 faces (see Figure 5.14). Between blocks, there was a 12 second rest period, before the participant was primed with a fixation cross for three seconds (which was followed by a 1s delay before the stimuli recommenced). To ensure the participants continued to attend the stimulus a one-back test was included. Each block had two repeats, in which the same image was shown twice in a row. The participants were required to press a button when they detected such a repetition. No cue was given indicating whether they were correct. A contrast between the faces and rooms was used to identify the FFA. The converse contrast (rooms−faces) localised both the RSP and the PPA.

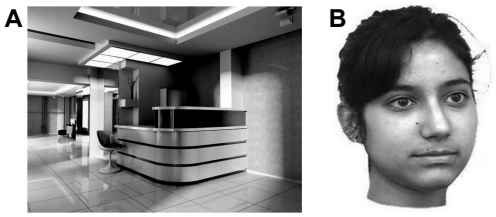


Figure 5.14: Example images from the PPA/FFA functional localiser. A. building interior. B. human face.

The functional data (after realignment) was smoothed with a 6mm FWHM Gaussian kernel. Using the SPM toolbox (SPM8, Wellcome Department of Imaging Neuroscience, London, UK) a univariate general linear model was applied to identify clusters responding more to faces or more to structures. Figure 5.15 illustrates the regressors used to localise these regions. Occasionally a sphere was used to aid in restricting the region of interest.

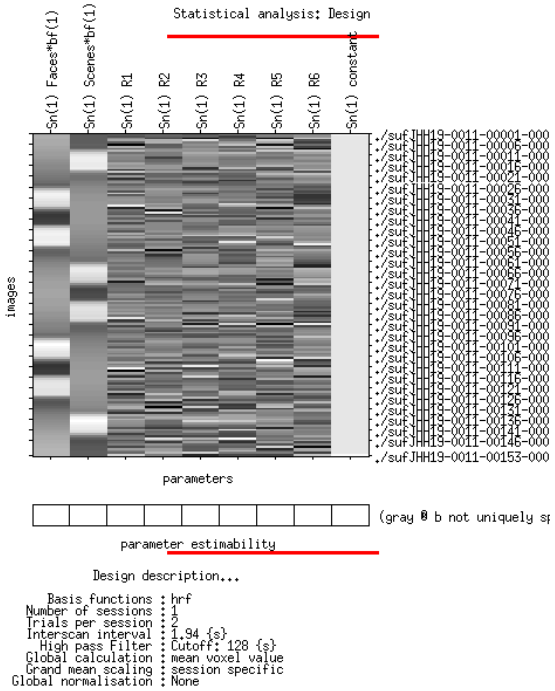


Figure 5.15: Example of the regressors used to analyse the functional localiser. The first two regressors are for the features of interest (faces or scenes being displayed). The next six regressors represent motion translation and rotation.

5.2.7 Preprocessing

fMRI data from the main experiment was also preprocessed using the SPM toolbox. As with the data in chapter 4, the fMRI volumes were first realigned to the first volume of the time series, to minimise motion artefacts. The data was then smoothed with a 3mm FWHM kernel. The images were kept in individual space, to avoid unnecessary processing. Slice timing correction was applied to minimise the effect of differences in slice times, especially important when interleaved and in an event related design.

5.2.8 Decoding

The default method for decoding was carried out in the following way. First the data was high-pass filtered with a frequency of $1/128Hz$. Rest periods were removed, and the data was z-scored; this meant subtracting the mean of each voxel time-course and dividing by the standard deviation of each time-course. The two or three volumes taken during each trial were averaged to help reduce noise and variation due to the haemodynamic response lag. The classification was performed through a cross-validation methodology. Each block was removed from the data, one at a time (leaving eleven remaining blocks). These remaining blocks were used to train the classifier, which was then tested on the left-out block. This is known as leave-one-out cross validation. The process was repeated twelve times, removing a different block each time. The mean classification accuracy was computed across these twelve classification runs. The classifier used varied, but by default we used a 4-way linear support vector machine classifier.

To approximately test whether the classification was significant, the trials were assumed to be independent and a binomial estimate was made as to the significance. For example 192 trials were performed for each participant. Given that there were four possible facing directions, by chance we would expect 48 of the classifications to be correct. The binomial distribution acting as the null distribution (i.e. assuming chance classification) predicts 59 trials or more would be correct 5% of the time (one-tailed) by chance. To perform a more rigorous multiple comparisons test, the classifier is retested on the same data but with the labels randomly shuffled. This is repeated 1000 times to provide an empirically derived null-distribution. The limited sample size means the accuracy of the distribution was quite poor. We fitted a normal distribution to this empirical estimated distribution to allow more precise estimates of the p-value at small probabilities.

Classifiers

We used the Princeton Multi-Voxel Pattern Analysis (MVPA) Toolbox to perform pattern classification on the fMRI data. Details of the classifiers and techniques are listed in detail in Appendix C.

5.2.9 Simple Decoding

As in the previous chapters, it seemed wise to confirm the method, algorithms, regressors and data were all accurate and functioning correctly. To this end, two simple classification tasks were developed. The most simple was to classify whether the participant was in the ‘mist’ (i.e. not navigating around the environment). The second was slightly more complex. It took advantage of our knowledge of the participant’s telemetry. The participants often spent several seconds rotating to face in the direction of the next target. During these turns, it was realised, the screen would contain global translating motion either left or right. The participants were not instructed to fixate, so it was uncertain the translating stimulus would have the desired effect. Decoding this condition was restricted to region hMT+, an area heavily involved in motion perception. Unlike in chapter 2, the region was localised anatomically, so

with considerable error.

The simple mist-no-mist classification task was (arbitrarily) decoded in the posterior cingulate (although many other areas would have success decoding this contrast). Indeed the posterior cingulate is not believed to be particularly relevant for visual stimuli. The regressors for each participant contained 192 samples for each condition (except one participant who only had 191 trials). To create the regressors, four volumes from the start of each manual control period were selected and four volumes from the period in the mist. As mentioned, these regressors were shifted backwards two TRs to take into account the delayed haemodynamic response.

To classify the direction of rotation, periods of at least two seconds of rotation greater than $30^{\circ}s^{-1}$ were identified. Regressors from the start of these periods lasting two TRs were created. To ensure the classification was not bias, the same number of trials were selected in each session.

5.3 Results

5.3.1 Behaviour

Participants, after training, were able to quickly and reliably travel to the correct location (from observation). Of more interest is the responses made in the pointing task. We recorded the accuracy and speed of the responses during the training and experiment. Figure 5.16 illustrates the number of trials which have a pointing error of more than 50°. The pointing error during the training improved considerably. Figures 5.17A and 5.17B illustrate the results for all participants for every trial. Note the concentration of responses during training about 0, 90 and 180 degrees is not due to a different strategy by participants, but rather is due to the use of a different (non-fMRI compatible) joystick during training. This joystick's movement range was restricted to a square, which meant the four corners (representing 0, 90 and 180 degree errors) were reached by moving the joystick to the limit of its movement. Intriguingly some participants (e.g. 1, 9 and 10) had concentrations at the start of training at the 45 and 135 degree angles, which meant they were tilting the joystick along N,S,E,W axes and not the diagonal ones. It appears these subjects learnt the directions were diagonal (NE,NW,SE,SW) before the end of the training.

All subjects except JHH19 (8) and YBR17 (10) achieved the threshold of getting 5 of the last 6 pointing queries correct during the training. These two subjects had one and two extra training blocks, respectively. By the end of these they had reached the required behavioural threshold.

During pilot trials it was found some cues were pointed to more quickly than others (figure 5.18A). It was unclear why this was happening, but the cues were repositioned in the diagonal directions, and the city and hills were replaced with a crane and a bridge. To confirm the four cues were responded to in equal time in the actual experiment an ANOVA was performed testing the combined times between cue responses. No significant difference was discovered ($F=0.61$, $p=0.61$). The cues were in different positions for different participants and so a similar statistical test was performed, by direction rather than by object ($F=0.61$, $p=0.61$). This may still lose the preference for particular objects or directions as the preference may vary between subjects. Figure 5.18B indicates that some subjects have unequal

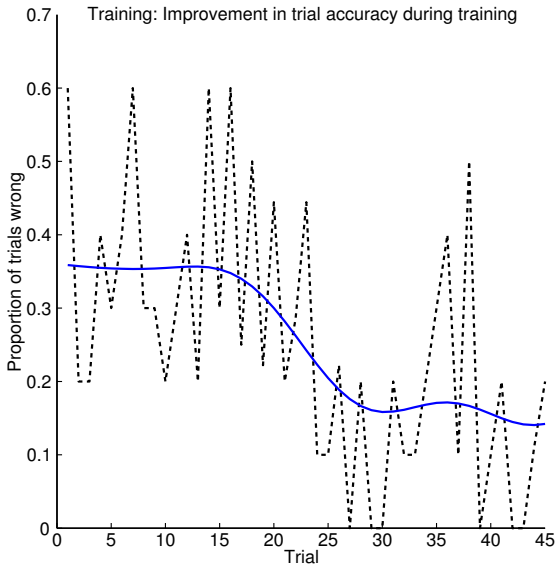


Figure 5.16: Improvement in number of wrong trials during training (averaged over all participants). The dotted black line indicates, trial-by-trial, the proportion of participants who got the answer wrong (over 50° error). The blue line is a Gaussian-smoothed curve (s.d. = 20 trials), which helps show the improvement during training. Participants do not start at chance (75% wrong) as the landmarks are visible throughout their time exploring the environment. An optimum participant could, in theory, get no trials wrong.

response times for the four directions. Once bonferroni multiple comparisons correction is applied subjects MLM23, YBR17 are still significant ($p<0.05$). It is intriguing that for the (uncorrected) significant participants the response times are made of two monotonic sequences (two up and two down)[2]. It could be one cue is being used as a reference and is the one being tracked during orientation changes, with the other cues are calculated using the orientation of this one. No particular cue appeared to be chosen across participants, when cue speed is plotted for individual participants (Figure 5.19A) and especially not when averaged across all subjects (Figure 5.19B).

5.3.2 Localising ROIs

Figure 5.20 shows, for one of the subjects, the localised regions of interest. The threshold used varied between participants. The value chosen was selected to ensure good coverage of the region, with minimal type 1 errors. Often a sphere was used to restrict the cluster to the region of interest (this avoided any clusters that were 'joined' being included). Appendix B briefly summarises the statistical strength of the localiser GLM tests used.

Besides the three functionally localised regions, four regions were localised anatomically. The posterior cingulate, thalamus, hippocampus and human middle temporal complex were identified using

[2]There is a $^1/_3$ chance of having such a sequence in four independent random numbers (determined through empirical testing), which means there is a 1.23% chance of four sequences having such an order. Although on the face of it this appears significant, this test is post-hoc, and is testing an unexpected pattern, detected after the experiment.

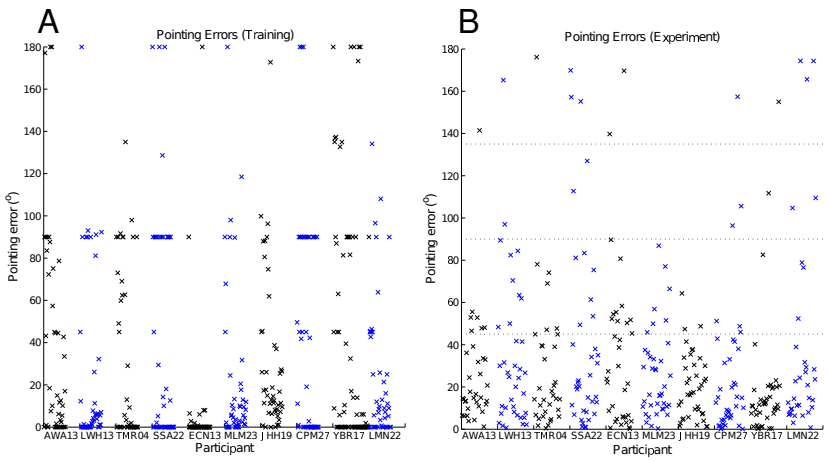


Figure 5.17: A. Pointing errors across subjects and trials during training. B. Pointing errors across subjects and trials during the experiment.

normalised anatomical masks from FSL's Harvard-Oxford cortical atlas [3]. Normalisation was calculated using SPM. The anatomical masks were converted to individual subject space by inverting the normalisation transform. Figure 5.21 illustrates the regions for one of the participants.

The Retrosplenial cortex is a subset of the posterior cingulate (Vann et al., 2009), but in the functional localiser it appears more posterior in many of the participants. Either the functional localiser is not identifying the RSP or the anatomical region defined does not extend rostral far enough. This localiser was based on Epstein et al. (2007). In their paper they note that their ROIs extended both superior (into area 23) and posterior (into the parietal-occipital sulcus). Appendix B has illustrations showing the discrepancy between the functional and the anatomical ROIs. As with the conclusion in Epstein et al. (2007) it seems reasonable to continue to use the functionally defined region as it is likely to include the region of interest. However, we will also proceed with the anatomically defined area, as insurance against poor functional localisation.

5.3.3 Classifying Simple Stimuli: Mist-no-mist

Figure 5.22A illustrates the subject-by-subject decoding accuracy for the mist/no-mist regressor in the posterior cingulate ROI, using the linear SVM classifier. Briefly, we also tested several other classifiers (see table C.1), on the same data. Figure 5.22B illustrates these results. There is a significant difference between classifiers on this data ($F = 6.87$, $p=3 \times 10^{-6}$). Due to the multiple comparisons correction however, the only difference which achieves significance is that all but nearest neighbour did significantly better than ridge regression. It seems fairly plausible however that Nearest Neighbour classifies this data worse than the other classifiers. Surprisingly LARS and LASSO did better on this data than the other classifiers. It was expected that, because the signal probably spans many voxels, less greedy clas-

[3] available from: http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases

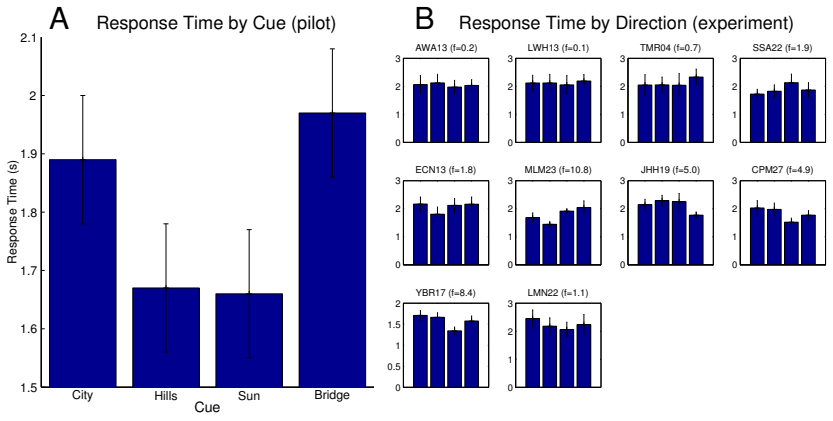


Figure 5.18: A. Pilot results: Pointing response times for the four cues, averaged across subjects. B. Experiment results: Pointing response times for the four directions, per subject. Error bars are 95% CIs.

sifiers (such as Logistic Regression) would do well. It is surprising how unsuccessful Ridge Regression was, although in retrospect it may be that a poor choice of preset penalty parameter was to blame. For other classifiers (e.g. PLS) the implementation set parameters by an inner cross-validation step on the training data. This data set was not the typical data one would normally classify, with a far more global correlated response than, for example, classifying motion or orientation direction.

Generalisable

It is important to note that these results cannot be generalised to classification of other stimuli. In particular the classification of place, direction or rotation will lead to far more subtle effects than those being classified here. Some classifiers (as discussed in appendix C) will be particularly suited to certain conditions. The purpose of this test was to confirm that the stimulus, MR imaging, timing, processing, regressors and classifiers were all functioning, and was not intended to be an exhaustive analysis of classifier performance.

A future study could investigate further by looking at the success of this set of classifiers in a variety of different conditions. These conditions should be chosen to allow the classifiers to be compared in response to the following three challenges: large number of voxels to small number of trials; large ROI with only a small clique of relevant voxels; correlated and uncorrelated voxel responses. These are probably the three key issues (besides the low SNR) which the classifier will need to respond to. This chapter is not specifically about classification methods, so such further work will need to be carried out in future studies.

Population size

One criticism of this study is the limited number of participants we were testing. However, the tests made here and in later results are not inter-group comparisons, but rather looking for a feature we expected

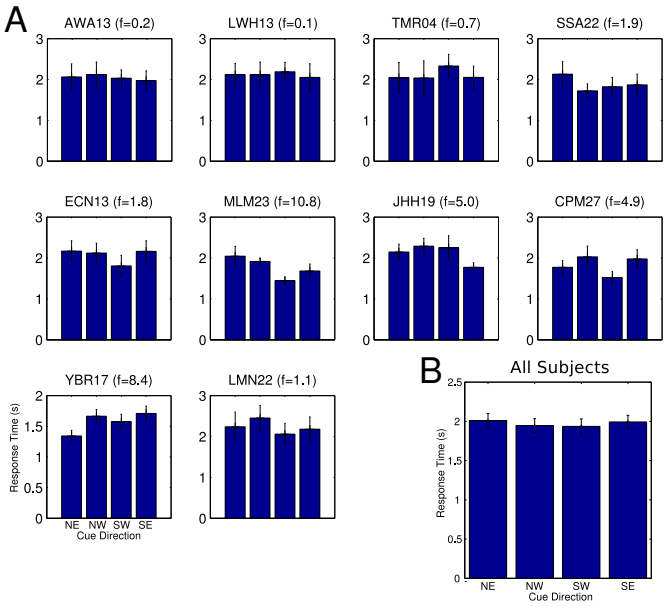


Figure 5.19: Pointing response times for the four cues, per subject and combined.

to find in all participants. This followed from earlier experiments such as that by Hassabis et al. (2009) in which only four participants were included, and (Rodriguez, 2010) in which eleven were used. Ten were included here as this seemed to offer a good compromise between unnecessary duplication and sufficient evidence. If a feature was detected in all ten individuals, one can surmise that it is likely to exist in the whole population the participants were selected from. Similarly, the group statistics (in which the classification scores were combined in a t-test) estimates the population mean and as a form of random-effects analysis is generalisable to the population.

5.3.4 Classifying Simple Stimuli: Direction of rotation

As a second test of the classifiers, regressor timing and methodology, we devised a new classification task. It was realised that the participants often spent several seconds at a time rotating in the virtual environment. These periods of rotation could be used to test the tool. As the participant rotates, the screen presents a global translation stimulus (either left or right). Region hMT+ is strongly associated with visual motion processing. The region was selected using an anatomical approximation and the direction of rotation was then classified using the voxel time courses from the region. Because the regressors were not predesigned, there was concern that they could suffer confounds. For example if the leftward rotations were concentrated during the earlier part of the experiment, long-term changes (e.g. induced by head motion) would allow a higher-than chance decoding accuracy, independent of neural activity. To avoid such confounds, first, the same number of left and rightward turns were identified for each session. Second, the decoding results in hMT+ were compared to the same classification in the posterior cingulate, an area we would expect not to be able to decode a translating visual field.

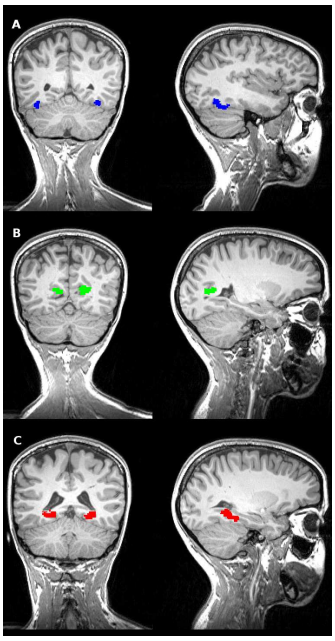


Figure 5.20: Example functionally defined regions of interest from subject JHH19. A. FFA, B. Retrosplenial cortex. C. PPA.

We found that we could decode the direction of rotation in hMT+ slightly better than chance (across subjects, one-sided t-test; p=0.042, t=1.95, mean=56.20, s.d.=10.08). In the posterior cingulate the classifier failed to decode the direction of motion (mean=50.64, s.d.=10.49. Compared to chance, p=0.43, t=0.19). However, the paired t-test comparing the results of these two regions did not quite reach significance (p=0.065, t=1.67, one-tailed t-test).

Within-subject decoding significance was tested using permutation testing for region hMT+. Of the ten participants five reached significance (uncorrected, 3 reached significance under bonferroni multiple comparisons correction). Figure 5.23 illustrates the p-values for the ten participants. Only 100 permutations were tested and the distribution assumed to be normal to allow a slightly more accurate estimate of the p-value (this was especially important for the three subjects in which none of the 100 permutations did better than the real classification).

In summary, we can not quite say that the evidence of classification is significant across subjects. However, there does appear to be a trend towards successful classification, a sign we have taken to indicate the classifiers are functioning correctly. This seems to be particularly visible in the single subject results.

Classifying Location

The place targets and task are very similar to the design of Hassabis et al. (2009), but with a pointing task incorporated. Our study also used more participants (10 instead of 4), hopefully allowing us to

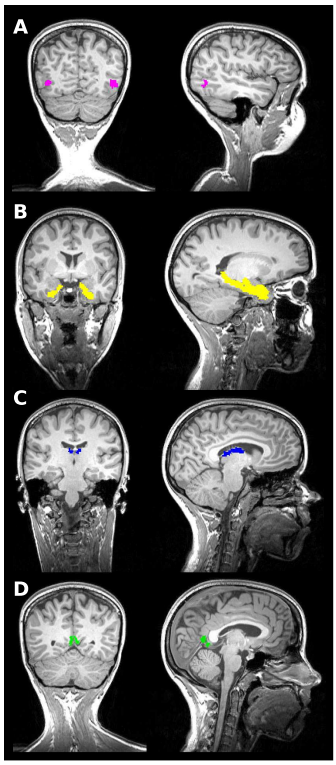


Figure 5.21: Example anatomically defined regions of interest from subject JHH19. A. hMT+, B. Hippocampus, C. Thalamus, D. Posterior Cingulate

confirm the results of the earlier study. However, decoding accuracies were very low, with little evidence of a significant classification effect. One of the first analyses did hint at some signal within the data. Figure 5.24 shows the decoding (of location) accuracies across participants. Assuming chance is 25%, the group's decoding accuracies were significantly above chance (classifying using PLS in the hippocampus, mean = 27.40, s.d. = 2.96, N = 10, p=0.015, t=2.56, one-sided t-test). These results were however invalidated using permutation testing, in which the classes of each trial were randomly permuted and the accuracy repeatedly sampled, to give an empirical null distribution. The PLS classifier required considerable computational time to solve, which meant sufficient numbers of permutations could not be calculated to estimate the null distribution for each participant. As a compromise, the average results of a limited number of permutations, across all participants was calculated instead. It was found to be slightly larger than naive-chance, at 26.42%. Note that this is significantly higher than one would expect (using the 66 permutations available). The standard deviation was also slightly larger than expected 4.2% rather than 3.2%. Repeating the above statistical test with this new chance-level gives a non-significant result of p=0.1612, which means there is no evidence that the data in the ROI contained place-related information.

Searchlight analysis was also applied to the hippocampal ROI. With between 2065 and 3051 voxels in each subject's hippocampal ROI, a considerable number of multiple comparisons will take place. A

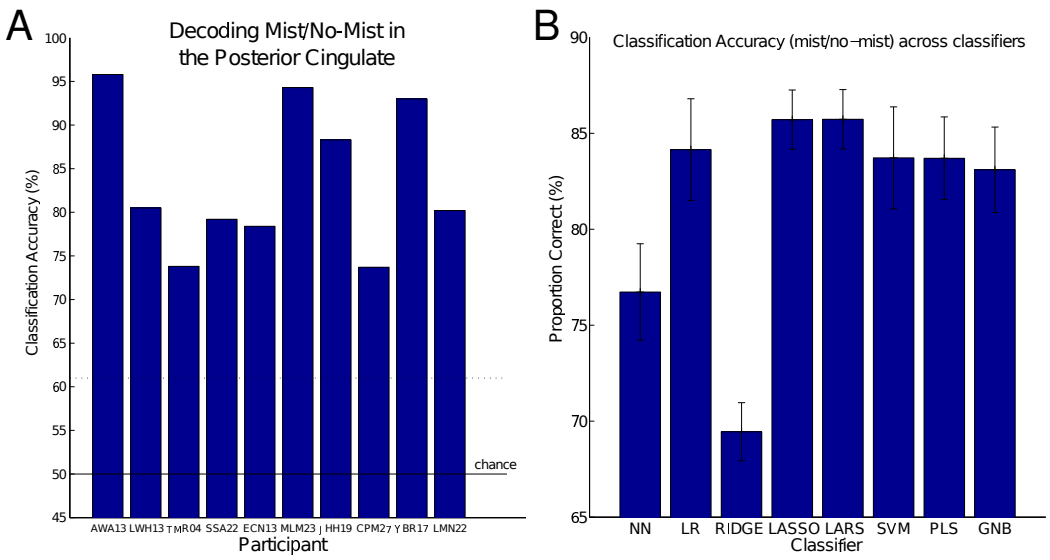


Figure 5.22: A. Accuracies when decoding mist/no mist in the posterior cingulate (lSVM). Dotted line indicates approximate $p < 0.001$ threshold. Solid line indicates chance performance. B. Classification accuracy of mist/no-mist in the posterior cingulate using different classifiers. Note, the ridge classifier may have had a poor choice of penalty parameter. Error bars indicate one standard error across participants.

large number of different combinations of classifiers, sphere sizes and parameters were tested but none produced a reliably significant result. Hassabis et al. (2009) did not directly correct for the number of voxels but instead compared between regions. This was also tested, but also found to not to be significant. A sizeable number of attempts were made, I've not recorded all these null p-values and results, but the multiple comparisons must also be taken into account.

Classifying Direction

No one has yet tried decoding the direction a participant is facing. There were several different regions which we hypothesised might contain head direction information; the posterior cingulate, the retrosplenial cortex and the thalamus. Table 5.2 summarises the decoding accuracies, note that several different configurations were tested. Oddly, running the configuration using only one trial of data found a significant (not multiple comparison corrected) result below chance for data from the retrosplenial cortex and the thalamus. Individual participants also seemed to often have significantly below chance decoding. For example, classifying LWH13 found only 29 out of 192 trials correct ($p < 0.00061$, which reaches one-tailed significance even when the 120 multiple comparisons are taken into account). Other researchers have also reported this (online and personal communication), but it is unclear how or why this happens. It has been suggested (personal communication) that such events are due to an unintended temporal correlations between trials of the same label.

As before, searchlight analyses were conducted on these regions. In general no significance was

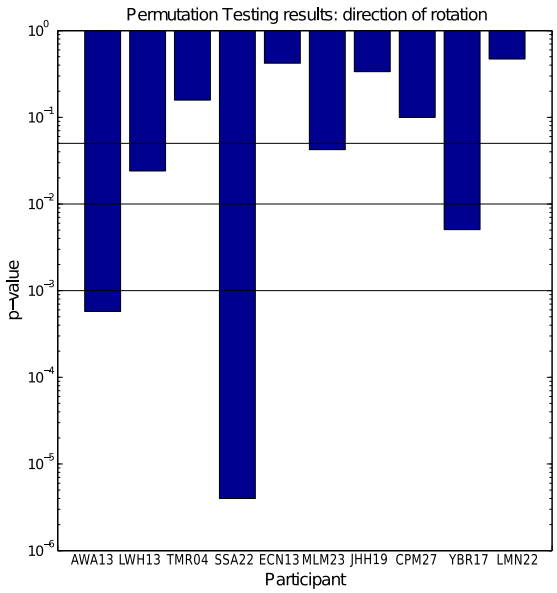


Figure 5.23: Classification of the direction of rotation, p-values, by subject. Decoding using the hMT+ ROI. p-values estimated using permutation testing.

found for any participant in any region. However, one participant (SSA22) had an intriguingly high classification accuracy in the thalamus (figure 5.25) with a similarly increased overall accuracy for that region. As this was restricted to one participant it seems likely it was chance however.

5.4 Discussion

5.4.1 Overview

Behaviour and Experiment

The participants all successfully learnt the environment and could all navigate successfully between locations and point at the distal landmarks. A possible pattern existed in their response times to different directions but it is unclear if this effect is real. Some participants said they kept track of the direction of a particular landmark feature (such as the archway). This could have meant that landmarks with a closer angle to that feature would have a faster response than those further away (maybe reflecting different response times when mentally rotating, e.g. Wraga et al., 2005). The environment and experimental design appeared to function well, although the failure to reproduce the place-decoding results suggests some features of the design need improving.

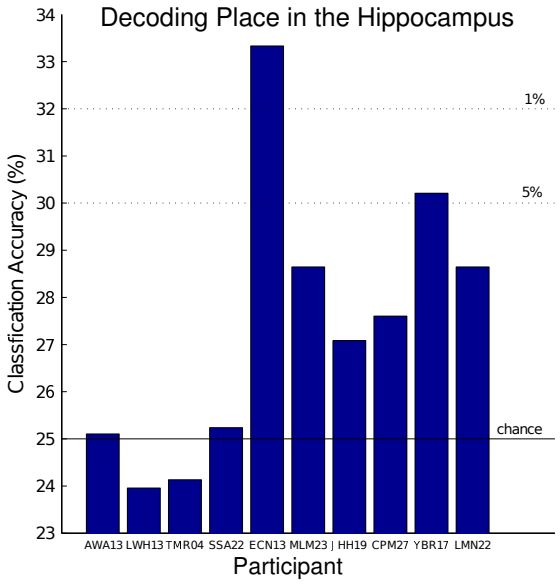


Figure 5.24: Decoding Location, accuracy across participants. Assumes chance is at 25% but permutation-testing suggests it may be higher (see text for details).

Localising ROIs

Most regions of interest were successfully localised using either anatomical toolboxes or functional imaging. The FFA was not identified successfully in two participants however. Of more concern was the localisation of the retrosplenial cortex. This was identified using a functional localiser and compared to the anatomical location of the posterior cingulate. A clear discrepancy appeared in many of the participants suggesting the functional localisation of the retrosplenial cortex was inaccurate. Without being able to be more specific or identify the cause of the discrepancy, both regions were used in classification.

Simple stimulus classification (mist and direction)

The simple stimuli were largely decoded successfully, which suggests that the regressors, scan, method, technique and analysis did not have critical flaws. In particular the presence or absence of mist was very significantly decoded. The direction of rotation had a weaker classification accuracy however. Five of the ten participants reached (uncorrected) significance, which suggests the stimuli and classification methods were functioning as expected.

Classifying Location and Direction

The head direction feature was not decoded reliably. This suggests that given this stimulus with this resolution fMRI scan and experimental design, it seems very uncertain whether the head direction of the participant can be decoded. Place classification was also very poor, with only the cross-subject t-test finding a significant result.

Configuration				Classification results (across subjects)		
Region	Classifier	Volumes	HDR delay (TRs)	Mean	Standard Deviation	p-value
Retrosplenial Cortex	SVM	2	2	24.4	3.50	0.58
	PLS	2	2	22.5	4.03	0.09
	PLS	2	1	22.6	3.08	0.04*
	PLS	3	2	24.0	4.89	0.55
Thalamus	SVM	2	2	25.8	3.43	0.47
	PLS	2	2	24.6	2.96	0.64
	PLS	2	1	23.4	1.64	0.012*
	PLS	3	2	24.3	3.60	0.56
Posterior Cingulate	SVM	2	2	26.4	3.25	0.21
	PLS	2	2	25.1	3.31	0.96
	PLS	2	1	22.8	4.85	0.19
	PLS	3	2	24.4	3.93	0.66

Table 5.2: Head Direction Classification Results. p-value is two-tailed (the uncorrected significant results are *below* chance performance.

5.4.2 Why did this study fail to classify location reliably compared to earlier place-decoding studies?

The simple stimulus classification analyses suggest that the tool and classifiers were functioning adequately. It is therefore likely to be something to do with the stimulus or fMRI imaging which has caused the failure to classify successfully. Why did the study by Rodriguez (2010) decode place with greater accuracy? It is impossible to say for sure, but there are several possible reasons which could explain the difference. The voxel size was considerably larger than in our study, $3mm \times 3mm \times 3mm$ compared to our $1.5mm \times 1.5mm \times 1.5mm$. This may have increased the signal/noise ratio (SNR) sufficiently for classification to take place. Although counter-intuitive, the SNR is reduced by a factor of $2\sqrt{2}$ for a doubling of the resolution (in two directions) (Constable et al., 1991). So (ignoring the slice thickness) our voxels will each have only $1/(2\sqrt{2})$ the SNR of the same voxels $3mm$ on a side. Constable et al. (1991) also note that recombining these voxels, in an attempt to regain the lost SNR will not completely achieve the increase in SNR. By combining four voxels one can see that by averaging the noise will be reduced by 50%, this reduction will mean the SNR will still be a factor of $\sqrt{2}$ worse than if the imaging had been done at the lower resolution initially. To summarise, it would seem that in the two later fMRI studies conducted during this PhD, imaging should have been performed at a slightly lower resolution (e.g. 2-3mm). Note that Constable et al. (1991) is concerned purely with anatomical MRI, with a related discussion being the question of spatial smoothing. Kamitani & Sawahata (2010) showed invertible spatial smoothing (e.g. Gaussian) did not affect classification (for certain types of filter). Spatial smoothing was used by Rodriguez (2010) prior to classification. Spatial smoothing may help reduce

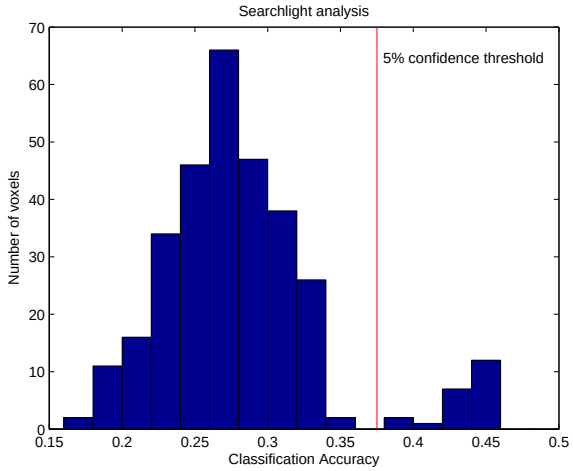


Figure 5.25: One participant's (SSA22) searchlight classification scores (from the thalamus). A small number of cliques appear to have a considerably greater classification accuracy. Dotted line indicates 5% (bonferroni multiple comparisons corrected) threshold.

noise, but it is unclear how useful it is. It should be noted that the effects of spatial smoothing are very different to downsampling (which can cause information loss). (Kamitani & Sawahata, 2010) also consider the effect of voxel size on classification and find that, assuming subvoxel representations are generating the signal decoded by MVPA, reducing the voxel's size does not necessarily provide extra information due to the increase in noise, cancelling out any benefit from smaller voxels. Hassabis et al. (2009) use $1.5mm^3$ voxels, a parameter we chose to mirror in our study. Their weaker decoding accuracies, compared to Rodriguez (2010), may be due to the SNR issue described above. A final step used by Rodriguez (2010) is pre-whitening using a GLM. This was tested on this experiment's data during earlier (unreported) classification attempts where the beta images from a GLM were used instead of the direct BOLD images. These are prewhitened during parameter estimation by the SPM toolkit.

Moving on to more experimental design aspects of the Rodriguez (2010) study. One very surprising aspect was that it was relatively brief (36 trials, 3 of each class). Such a small number of training samples (compared to 192 in this study) would be expected to produce poor classification results. However, the brevity of the experiment may help protect participants from overlearning or fatigue. Also extending the length of each trial improves the SNR of each trial's BOLD signal. In their study, the pretraining occurred one or two days prior to scanning. Speculating a little, could some form of reorganisation (through hippocampal replay) or consolidation be occurring in the interim? To be relevant for decoding, more cells would have to be recruited to represent the environment during replay. Although an intriguing idea, place fields appear to become very quickly fixed and then remain reliably placed over weeks in multiunit recording experiments (in rats). Hassabis et al. (2009) trained participants immediately prior to scanning (as in our experiment) and the accuracy of decoding was also quite limited in their experiment (compared to the more recent Rodriguez (2010) study).

The behaviour of participants in the Rodriguez (2010) study was quite intriguing. The reached the target on average only 67.4% of trials - indicating considerably error and difficulty. This means the participants were probably concentrating and attending the task far more than in my study or that by Hassabis et al. (2009). Especially towards the end of my experiment it was noted participants became fatigued. Could this explain the difference in decoding accuracy? To test this hypothesis, only blocks from the first session were used in the cross-validation. No benefit was found. Participants were asked about their strategies, and all appeared to still consider the task a spatial one. A few imagined a map of the environment and several used particular landmarks throughout which they kept track of to remain orientated (the arch and the front of the house).

A final obvious difference was the task - in our experiment the participants were required to concentrate on an orientation task, rather than the place task. It might be that this distraction had a weakening effect on the place signal in the hippocampus. This is especially true when one realises that, at least during the mist, the participant's *location* was no long of interest to them (in completing the task), they merely had to recall their orientation. In Hassabis et al. (2009) the task did not require them to recall their location while looking downward, while the task in Rodriguez (2010) seemed to require considerable, continuous processing of location by the participant.

In electrophysiological experiments conducted by Hok et al. (2005), place cells were found to congregate around goal locations. In all three fMRI studies (this study and the two cited above), particular goal locations are highlighted. Maybe the hippocampus is responding more strongly to being at or very near the goal location. However, any experiment which tests this must be careful to avoid the confounded caused by reward and learning at the goal location. The study by Spiers & Maguire (2007) investigates this, using the convoluted one-way system of inner London to disassociate distance to the goal location from time-to-goal. They didn't find correlated activity in the hippocampus itself, but the subiculum was found to have a strong *negative* correlation with the distance to the goal, an intriguing but unresolved finding. How these findings might explain or assist with the decoding is unclear, but a differential increase in BOLD signal when approaching different goals may be one method that could explain the results, without requiring subvoxel anisotropies in place cell distribution.

Our study was designed to carefully avoid visual confounds for either place or direction. Hassabis et al. (2009) also somewhat achieve this by making the visual field rotate downwards. We went one step further and not only masked the visual scene but ensured that the visual cues would not provide location or orientation information in the majority of conditions thus avoiding any 'visual memory'. The cue card in Rodriguez (2010) may have provided such a cue, although they found that in their approximately localised V1 ROI there was no significant decoding of place. The presentation of the cue card may still have been sufficient to involve the hippocampus in scene pattern-separation. This may seem implausible due to the sparse environment, but a recent study (Bonnici et al., 2011) has showed that very small differences in a visual scene (potentially invisible when decoding in V1) are successfully decoded from a hippocampal fMRI ROI due to its particular pattern separation abilities. Future studies investigating place decoding must therefore be vigilant for visual confounds. Their complete avoidance may not be possible, as the visual scene (or recent memory of the visual scene) are intrinsic parts of the

navigation task.

5.4.3 Decoding Head Direction

Although of poor significance, there appeared to be indications that place was being classified above chance from the hippocampus. Similarly the simple-stimuli were decoded successfully. The head direction signal conversely gave no indication in any way of classification above chance in any of the three regions (retrosplenial cortex, thalamus, posterior cingulate), with the only exception being a small clique of voxels in the thalamus in one participant. This absence of any signal may be due to a failure in the design of the experiment, but it is likely to be due to more fundamental differences in the organisation of the regions being investigated. In particular it may indicate that head direction cells are not functionally clustered. If they are organised in a homogeneous manner, then the BOLD signal from the voxels in that region will all have a similar response to any head direction. At first it seems place cells also appear are evenly spread. However, the electrophysiological recording study by Eichenbaum et al. (1989) found strong evidence that nearby place cells are more likely to have overlapping place fields than distant cells. This strongly suggests an inhomogeneous organisation of place fields in CA1, which would provide at least the possibility of decoding using fMRI.

In the regions containing head direction cells, no similar organisation has been reported. For example, in the postsubiculum Taube et al. (1990) found that there was no systematic shift in the peak-firing direction as the recording electrode was moved obliquely through the cell layers. This lack of organisation makes it much less likely to be successfully decoded using fMRI MVPA than place cells in the hippocampus.

5.4.4 Future Research

Understanding why the study by Rodriguez (2010) achieved far higher classification accuracies than Hassabis et al. (2009) or this study would allow future experiments in this field to be performed more reliably. Although the differences may be explained by differences in voxel size, it would be interesting to test for the contribution of visual cues to hippocampal decoding.

One aspect that was not investigated in this study was the relative contributions to the classification of different hippocampal subfields. Had the searchlight classification been successful this would have been investigated, by rendering searchlight results on the anatomical image. Recent advances in segmentation methods now allow such precise distinctions to be made (Bonnici et al., 2012), although this segmentation must still be performed manually.

Regarding the head direction component of the study, the design of the experiment could be modified to increase the signal strength. An obvious change would be to increase the size of the environment and encourage long periods of motion in a single direction. This would allow a block design to be used, with the associated increase in BOLD signal. Such a design is already used by studies investigating grid-cells using fMRI (Doeller et al., 2010).

For decoding to take place there must be some form of anisotropy in the organisation of cells. In the hippocampus this might be achieved by sparse coding and pattern-separation. For head direction cells however, there have not been any findings showing a systematic organisation of cells with similar direction preferences (e.g. Taube et al., 1990). This suggests that head direction cell organisation may be isotropic, making decoding of the person's direction (using head direction cells) exceedingly difficult, if not impossible. One might consider adaptation analysis instead, but it has been regularly reported that head direction cells do not experience adaptation in their firing rate. However, Taube et al. (1990) did note that a few cells only fire a dozen or so spikes, as the head direction rotates past their preferred direction. This was reported to be independent of the time spent facing their preferred direction. One hypothesis for future fMRI studies then, is that regions containing such cells will experience a higher BOLD signal during periods of faster rotation than slower, as more of these cells will be 'triggered' per second, during the fast rotation.

In summary, fMRI and MVPA seem to be well placed to begin investigating hippocampal activity in humans. However, care must be taken to ensure that the method can distinguish between place-related activity and the visual, memory and recall related tasks the hippocampus is also engaged with.

Chapter 6

Conclusion

Overview

This thesis has examined the psychophysical and neurological underpinnings of our perception of motion and self-motion. In Chapter 2 the research focused on the columnar organisation of the cortex and the response to purely unimodal stimuli. It found evidence for columns in hMT+, a large-scale map in V1 and a preference for motion in vertical directions in both the strength of the BOLD response and in the psychophysical accuracy of the participants. It also analysed the results of an optical imaging experiment which appears to show that columns do have a response which supports the chapter's hypothesis.

Chapter 3 extended the analysis of motion into the integration of more naturalistic self-motion stimuli with both auditory and visual self-motion cues. The integration of these cues closely followed the predictions of the Bayesian model. A mechanistic model which used divisive normalisation was modified to allow it to produce predictions for the psychophysics. It was found to produce similar results to the Bayesian model. The two models were then modified to fit the conflict condition's results. These modified models were still able to predict the congruent condition's results. The experiment also investigated the psychophysical adaptation effect, in which motion in an opposite direction is perceived after presentation of a motion stimulus. The adaptation effect appeared to vary between the stimulus types, and depended on the actual stimuli and not the perception of the stimuli. This offered suggestions as to where in the cortical hierarchy the adaptation was taking place; namely it seemed it was mainly localised in unimodal regions.

Clearly, we wanted to find out more about the neural underpinnings of these effects. In particular we wanted to test whether the integration of the cues was performed in the ventral intraparietal sulcus, as had previously been hypothesised, and whether the reweighting was represented by changes in connectivity between unimodal regions and the putative integration region. We used several methods in Chapter 4 to investigate these two questions, including Dynamic Causal Modelling (DCM) and Structural Equation Modelling (SEM). DCM did suggest that the model incorporating coherence related changes in connectivity. However the direction of the effect was inconsistent across participants, casting doubt on

this conclusion. Adaptation, super-additivity and cross-modal classification were used to test whether VIP was a true integration region. In spite of the wide range of tests the experiment was designed to perform, the results were very equivocal. It seemed the underlying problem was the limited BOLD signal produced by the stimuli. Several suggestions for improvements have been made, a block-design, larger voxels, rotating self-motion stimuli and higher contrast stimuli. With these changes a repeat experiment should be considerably more successful.

Chapter 5 concludes the research by investigating whether pattern classification can be applied to head direction regions. An obvious extension of both the above research and the earlier studies by Hassabis et al. (2009) and Rodriguez (2010) was to look for successful classification of the head direction signal in regions such as the the retrosplenial cortex and the thalamus. Place was decoded to a level that was just significant, while other features, including the rotation direction of the participant (from visual areas) were also decoded. The head direction signal gave no indication of being detectable. It was hypothesised that this was because head direction cells lack any spatial organisation, compared to place cells which may have a slight organisation.

Chapter 2: Simple Planar Motion

Chapter 2 investigated the visual motion regions hMT+ and V1 and the regions' responses to motion in the four cardinal directions. A method for possibly detecting columnar organisation was developed and tested on the two regions. A large and small scale organisation was found, in V1 and hMT+ respectively. Specifically, hMT+ voxels were found to have greater responses to opposing directions of visual motion, while V1 voxels seemed largely agnostic regarding the direction of the visual stimulus. Conversely, a large-scale organisation in V1 matched both the known retinotopy of the area and the anisotropic response of the region to planar visual motion. The combining of multiple subject maps to maximise signal, using the freesurfer tool, allowed this anisotropic organisation of V1 to be clearly visualised. A final fMRI result was the finding that the mean voxel t-value across all voxels and participants was significantly greater in the vertical axis than in the horizontal. Corroborating this last result was the unexpected supporting evidence from both the participants' response accuracies and the results of psychophysics studies. In particular the result that vertical motion speed differences are easier to distinguish appears to fit with our finding that vertical motion is more strongly represented in V1. The chapter's main result, that the method described can detect columnar organisation, fits nicely with the analysis of the results of Kaskan et al. (2010)'s optical imaging study. The brief analysis found a strong negative correlation between orthogonal responses at scales of about $400\mu m$. Such an organisation, it could be imagined, would bias a voxel to respond less to orthogonal visual motion. The amount of bias would vary depending on the size of the voxel and the size of the columns, possibly opening an avenue to investigate the scale of columns in human cortex. This would require a much larger dataset of optical imaging maps. This dataset could then be used in a computational model to allow voxel sampling of the cortex to be predicted. The point spread function (PSF) of fMRI needs to be taken into account in these models. Different levels of correlation should be expected, depending on the ratio of the column width

to the PSF width (similar to the results of figure 6 in Yacoub et al., 2008). By changing the PSF the change in correlation at different scales can be estimated. It would then be a matter of fitting the curve of correlation values to the modelled prediction. The experiment needs to be repeated with a retinotopic mapping and hMT+ subfield identification, preferably in a high-field MRI scanner. This would allow the correlated voxels to be more precisely localised and associated with a particular subfield.

Measures of columnar organisation and integrity may be applicable within clinical work. Differences in columnar organisation appear to be a feature found in several clinical groups. For example mini-columns appear to be smaller and more numerous in those with Asperger's syndrome (Casanova et al., 2002). It would be interesting to see if the correlation results differ significantly between this group and controls. Of greater clinical relevance is the possible link with columnar-scale neurofibrillary tangles (NFT) (Esiri & Chance, 2006) associated with Alzheimer's Disease. Hoesen & Solodkin (1994) concludes that 'a growing body of evidence suggests that AD pathology attacks at least some aspects of the modular organization of the cortex'. It seems reasonable to hypothesise that changes in columnar organisation may occur during ageing or Alzheimer's Disease. Looking for such changes using the method outlined in chapter 2 may provide a clinical measure of disease progression or type.

Chapter 3: The Psychophysics of Audiovisual Selfmotion Integration

The accurate perception of one's self-motion in the environment is clearly vital for survival. The planar motion stimuli in chapter 2 are highly simplified examples of stimuli one might experience by moving (for example, when rotating around a vertical axis). It is through the combining of different cues that one can achieve much more precision in estimates of self-motion. In chapter 3 it was decided that visual and auditory self-motion cues would be used to simulate self-motion in a forward direction. Sessions in which only one stimulus was present were used to estimate coherence thresholds for each participant. These are the levels of coherence in the two modalities for which participants could achieve a certain accuracy. In some sessions both stimuli were presented in a congruent manner. In a final session they were presented in conflict.

The experiment was designed to investigate how the unimodal and bimodal thresholds were related. In particular several models were proposed to predict the bimodal results, given the unimodal values. The most obvious model was Bayesian integration. This predicts that the two modalities would be integrated by multiplying the two probability distributions to create a new distribution with a mean weighted towards the unimodal distribution of least variance. It was with little surprise that this model was found to fit the data very well. A naïve model, for comparison, assumed that no integration took place, and the bimodal results would reflect the most coherent unimodal thresholds. This was found to be a very poor fit to the data. A model proposed by Ohshiro et al. (2011) uses a simple divisive normalisation function to predict how a layer of neurons could integrate the signals from two unimodal layers. The model had to be adapted to function in the psychophysics domain - modifications not yet verified in other experiments. In particular the changes include noise within unimodal neurons, a modification future studies could incorporate.

Adaptation appeared to have a large influence on the results of the experiment, with the presentation of a stimulus leading in one direction leading to an increase by between 7.8% to 12.5% in following trials declared to be in the opposing direction. It was found that this adaptation appeared to vary between modalities (the visual and bimodal conditions having greater adaptation than the auditory condition). The effect of repeated trials was also investigated and it was found that only in the visual condition did additional trials introduce greater adaptation. These results pointed to the conclusion that the adaptation was localised to the unimodal regions. To test this, the normalisation model was altered to incorporate the adaptation as a modification to the neural model. The adaptation was constrained to a subset of the three layers. The strength of the adaptation in the three locations was adjusted to fit the unimodal adaptation results, then the bimodal adaptation was tested. It was found that the model fitted the experimental results most accurately when adaptation was constrained to the unimodal regions alone.

A final line of evidence tested whether the adaptation was to the actual stimuli or to the perceived stimuli. It was found that the actual stimuli resulted in far greater adaptation than the responses of the participants. This would suggest the adaptation is at an earlier stage than any decision making layer.

Although these lines of evidence all point towards adaptation being restricted to unimodal regions, the use of cross-modal adaptation experiments would be far more capable of distinguishing where the adaptation is localised. For example a leftward visual trial followed by a leftward auditory trial should not experience adaptation, if the effect is restricted to unimodal regions.

Chapter 4: Self-motion: An fMRI study

One of the motivations for the visual stimulus of chapter three and four was the finding in microelectrode recording experiments that the ventral intraparietal sulcus contains centre-of-expansion location selective cells. The fMRI study tested whether the region does combine the two modalities by looking for superadditivity, cross-modal adaptation and cross-modal classification. For superadditivity, the hypothesis was that the bimodal stimulus condition would elicit a greater response than the sum of the two unimodal conditions. For cross-modal adaptation it was predicted that a leftward stimulus in one condition would cause greater adaptation to leftward motion in the alternative stimulus than rightward motion, thus indicating that the same neurons are responding to motion in either modality. This assumed that neurons would respond to congruent pairs of motion (i.e. a neuron which responds to leftward visual selfmotion would also respond to leftward auditory selfmotion) - an assumption in retrospect that may be invalid. For example, Bremmer et al. (2002b) found that most visual-vestibular bimodal neurons in VIP responded to noncomplementary motion; a visual dot-motion field moving rightward would trigger a cell which responds to rightward physical rotation. A complementary response would respond to the vestibular motion which matches the visual stimulus one would experience if rotated. Both adaptation and classification were inconclusive on the unimodal stimuli; there was no significant adaptation between directions, however the type of stimulus did invoke adaptation. In particular region VIP was found to experience adaptation in visual trials which have followed auditory trials. This result may be due to other effects, for example changes in the participant's level of attention.

The experiment was also designed to test the hypothesis that the connections to the integration region are modulated by the coherence of the stimuli. To do this, each block consisted of an audiovisual stimulus in which either the visual or auditory stimulus had a greater coherence. Dynamic Causal Modelling (DCM) and Structural Equation Modelling (SEM) were both used to test whether significant changes in connectivity occur between the three regions.

The DCM analysis reported a significant result across participants, using the random effects (RFX) analysis, that our alternative model which includes modulated afferent connections was more likely than the model with fixed-strength connections. The modulation is controlled by which stimulus is more coherent in that block. For each participant a pair of parameters for the two connections, weights the strength of the modulation. Although this model was found to be significantly more likely than the unmodulated model, it was found that the direction of the modulation across participants varied. This suggests the effect is, at best, not what we were expecting, and at worse a flaw with the DCM analysis. Poor model fit is clearly a potential issue. In summary, even though one model was found to be significantly more likely than another, *both* models may be of very poor fit and so lead to invalid inference about the underlying connectivity. However, Stephan et al. (2010) and others have suggested that a poor model fit will not lead to a false inference. The extra parameters in the second model means that the prior distribution is spread across a much larger space, which means the cost of the extra parameters, if the null hypothesis is true, would lead to the modulated model being considered more unlikely than the unmodulated model. Such issues could be tested by repeating the DCM tests on randomly permuted data, such that the coherence of the two modalities is no longer correctly assigned in the model. This, one would expect, would lead to the unmodulated model becoming significantly more likely than the alternative. The much less complex structural equation modelling tool was also used. A model was fitted to the data in each of the two coherence conditions. There was no trend across participants for connectivity changes between the two models.

The poor results in all the analyses, including simple BOLD signal strength analyses, suggest there was an underlying problem with the limited BOLD signal produced by the stimuli. To improve the experiment, obvious changes include switching from an event-related to a block-design. This usually improves the signal strength and makes various tasks such as classification and adaptation analysis far easier. By increasing the voxel size one should also be able to improve the signal-to-noise ratio. The current stimulus, although designed to be a clear self-motion stimulus does not seem to be easy to classify and would be challenging to extend in time, into a block design. The stimulus could instead be a rotation around the dorsal-ventral axis, as in Bremmer et al. (2002b). VIP neurons respond to such stimuli, so the experiment should still be able to investigate the same hypotheses.

Chapter 5: Head Direction

The direction of motion relative to the head or eye direction is an example of an egocentric feature; the allocentric direction was not believed to influence the responses of VIP neurons or the responses in hMT+ or A1. To extend the research, a final experiment was conducted to investigate the head

direction in allocentric space. In particular it is known that various well conserved regions of the brain process allocentric information, for example place, grid and head direction cells have all been recorded from several species. These are all examples of allocentric responses: They depend on the subject's orientation in the environment. The obvious next step, beyond the fMRI experiment of chapter 4 was to look for head direction correlates in allocentric coordinates rather than just in egocentric coordinates. Although apparently a similar topic, the task and the brain regions involved are quite different. The task required the participant to become familiar with a 3d environment and navigate within the environment to different locations and face different directions. The brain regions focused upon were those in which head direction cells were found during multiunit recordings in non-human, for example the retrosplenial cortex.

The experiment was developed with two earlier fMRI studies in mind, in which the subject's location was decoded from a hippocampal ROI. Hassabis et al. (2009) and Rodriguez (2010) were able to successful classify which of four or three locations the participant had visited in a virtual environment. The experiment followed closely the design described by Hassabis et al. (2009), but with an added rotation at each location to allow the testing of head-direction as well as place.

Other stimuli were classified successfully; the direction the person was rotating and whether they were in the mist-period. Importantly the location of the participant was (just) classified above chance, across subjects. However, the head direction of the participants remained well below any significance. There was no indication that head direction could be classified from any of the three regions of interest. This absence of classification could be due to a poor BOLD signal (e.g. due to small voxels, event-related design, weak attention or insufficient samples), however, the organisation of the underlying neurons may be the explanation. Place cells have been reported to cluster with others having similar place fields (Eichenbaum et al., 1989). No such organisation has been reported for head direction cells. Without such inhomogeneity at any scale, the response of a voxel will be the same regardless of the participant's head direction.

The place-classification although apparently successful, was very weak compared to the results of the earlier studies. It may be that the efforts in our study to avoid visual confounds are masking the effect Rodriguez (2010) found so clearly. It has been suggested that one can 'remember' the previous scene even if the visual stimulus is masked. This memory might be an important driver of activity in the hippocampus. The landmark cues within the environment were carefully placed so that they (mostly) would be invisible, or at least different, in the sixteen location-direction combinations, in an attempt to avoid this confound. An experiment in which visiting some locations results in identical visual scenes, while others result in differing visual scenes could be created which might help test the importance of such visual features, and whether their presence can explain the difference between this and earlier experiments.

Summary

The experiments in chapters 4 and 5 were relatively ‘high-risk’, with limited evidence in the literature for some of the hypotheses. They both relied on detecting small changes in the BOLD signal against a background of considerable noise and variance. Many suggestions have been made for future studies to address the same questions. In particular the self-motion fMRI study of chapter 4 could be repeated with different stimuli. Chapter 2 describes a method for investigating the columnar organisation of the cortex using fMRI. Considerable work is needed to expand and refine the method. Similarly, in chapter 3 the use of the Normalisation model to predict psychophysical results offered some intriguing hints at differences in its predictions compared to the descriptive Bayesian model. Finding experimental protocols which separate these models may allow the mechanisms underlying the integration process to be partly determined through psychophysical study.

In summary, this thesis has investigated both the perception and processing of self motion cues, separately and in the way they are integrated.

The experiments and their results have led to several possible avenues for future research, which I hope will soon be followed.

Appendix A

Self-motion fMRI study: Localising

A.1 Background

There were several regions we wanted to localise. Initially the most important were the Ventral Intraparietal Sulcus, hMT+ (and its subfields) and the Planum Temporale. Other regions of interest include V1, which also encodes visual motion stimuli and A1 which has been reported to represent auditory motion stimuli. Below, methods to localise VIP and hMT+ (and subfields) are described.

A.1.1 VIP

Several groups have reported localising VIP using a variety of stimuli in fMRI experiments. Below we summarise these localisers, in an attempt to select the most successful. Of particular concern was the scan time required. We wanted the localiser to take no more than 15 minutes, so we could dedicate the majority of the experiment to the stimuli of interest. Below are three methods for localising VIP.

Saccades and Smooth Pursuit

Konen & Kastner (2008) used a protocol in which the subjects saccade in different directions. This produced topographic maps in each of the subfields along the IPS allowing them to be distinguished (referred to in their paper as IPS1-5). Of interest was the finding that smooth pursuit eye movements (SPEMs) were preferred in some subfields more than others. They report that those patches of activation in response to SPEMs were most often found in IPS3-IPS5 (IPS5 being the likely candidate for VIP). They note that this fits with the electrophysiology literature in which the majority of VIP neurons respond to SPEMs (Schlack et al., 2003). By combining IPS5's lower response to the saccades and its greater response to SPEMs one might be able to use this method to localise IPS5/VIP.

Retinotopic Mapping

Several studies have used polar angle mapping to discover the boundaries of areas within the posterior parietal cortex (including those within the IPS) (Sereno et al., 2001; Schluppeck et al., 2005; Silver et al., 2005). Sereno & Huang (2006) and Konen & Kastner (2008) have been able to identify a region IPS5, the putative ventral intraparietal area (VIP). The study by Sereno & Huang (2006) was also able to localise the region by somatosensory stimulation using air-puffs on the face applied at twelve different locations. Both of these methods however appear to take considerable time (an hour of scan time) and do not appear to be reliable across subjects.

Selfmotion stimuli

More recently there have been attempts at localising the putative area VIP using its strong response for self-motion cues. Wall & Smith (2008) found two regions of the brain, VIP and the cingulate sulcus visual area (CSv), which both responded far more strongly to a self-motion stimulus containing a single global flow than to a stimulus in which the global stimulus motion was disrupted. They suggest that other regions (such as MT) would respond equally to both stimulus types, while those responsible for encoding self-motion cues will respond more to the self-motion compatible stimulus.

The stimulus used was a field of expanding, contracting and rotating dots, which appeared to spiral in and out of the screen, sometimes rotating one way, and sometimes the other (Figure 4.10A). This changing stimulus was chosen because earlier work (Morrone et al., 2000) looking at the division of hMT+ found that such variation in motion elicits the strongest response. The incompatible stimulus was made of nine smaller tiles, each with the same stimulus as above (Figure 4.10B). The nine separate centres of expansion are not compatible with real self motion, and so shouldn't cause as much of a response in regions which respond specifically to self-motion cues.

A similar, but more extensive study by Cardin & Smith (2010) had similar results, but found seven regions which had a greater response to the ego-motion compatible stimuli. Besides hMT+, VIP and CSv the study found four other regions: Parieto-occipital sulcus (pV6), posterior Insula (PIVC), the dorsal portion of the postcentral gyrus (putative 2v, p2v) and the precuneus (Pc).

Because of the proximity of p2v, there was concern this area (or other regions not yet identified) might be mistaken for VIP. To minimise this the talairach coordinates for each cluster were compared with those from both Cardin & Smith (2010) and Wall & Smith (2008).

Cardin & Smith (2010) report that VIP was found in 86% of the hemispheres (at $p < 0.005$ uncorrected), using data collected during 29 minutes of scanning.

As other localisers were also needed, there was concern that not enough data would be collected in the limited time available for localising the region. However, this appeared to be the most reliable localiser available. It was this stimulus which we decided to use in our localiser.

MT/MST

Experiments, such as those conducted by Zeki et al. (1991), began the revolution in neuroimaging investigating the functional response of brain regions to stimuli. Zeki et al. (1991) used a stimulus consisting of moving black squares and used Positron Emission Tomography (PET) to measure cerebral blood flow. They also made early use of statistical parametric mapping to perform multiple comparison correction and the display of the experiment results. They found a very strong response in a region in the temporal lobe (V5), which appeared to match the motion complex in monkey (MT+).

The area has since been extensively studied, with various attempts at distinguishing the subfields of hMT+ using fMRI. Several of the methods used to distinguish the subfields are outlined below:

Pursuit Eye Movements

Dukelow et al. (2001) used a visual pursuit experiment in their study to distinguish MT and MST. Subjects tracked a white dot which moved slowly around the screen. This experiment was based on the electrophysiological research of Newsome et al. (1988) in which single cell recording revealed strong correlations between the movement of the eye in its pursuit of the target, and the firing rate of cells in MSTd and MSTl. The pursuit task in humans generated an increase in the BOLD response in putative MST.

Retinotopy

In looking for differences in MT's and MST's responses, Huk et al. (2002) cites prior findings that suggest MT has a much more coherent retinotopy than MST. They go on to hypothesise that MT will respond much more strongly to a retinotopic mapping stimulus. A rotating wedge was used, with those voxels experiencing the greatest modulation to the wedge's motion classed as MT. This aspect of their experiment involved between 17 to 26 minutes of scanning. Yan & Wu (2010) also report using this method in their conference paper, as a way of identifying area MT. Although the method had some success, the long scan time deterred us from using this localiser.

Ipsilateral Receptive Fields

Electrophysiological experiments in animals have found that MST has much larger receptive fields than MT. For example, Desimone & Ungerleider (1986) report field sizes approximately 75% larger in MST than in MT, for the same eccentricity. One effect of MST's large receptive fields is that, unlike MT's receptive fields which are generally restricted to just the contralateral hemifield, MST's extend far into the ipsilateral hemifield. By presenting motion stimuli in only the ipsilateral hemifield it was predicted by Dukelow et al. (2001) that only the MST subfield of hMT+ would respond. Their analysis did appear to distinguish the two regions, with hMT more posterior to hMST. However, the regions identified appeared quite disconnected and unclear. The total fMRI scan time for this part of the experiment was at least 15 minutes for each participant.

Huk et al. (2002) also looked at a variety of stimuli to localise MT and MST. One part of their experiment also used the differences in ipsilateral receptive field extent. In four of the five participants they reported that they were able to separate MT and MST using this difference. Again, looking at the figures they provide; although there is an obvious division in the areas responding to the stimuli, it is far from certain that this method would provide sufficient accuracy or volume for use as a localiser for further experimentation. The total scan time for this stimulus contrast was between 25 and 50 minutes.

This method was used, by Wall & Smith (2008) to localise MT and MST. In this case this localiser was only run for 8 minutes in a blocked design. Recently it was used by Fischer et al. (2011) in which the relevant stimuli for this contrast were presented for a total of 15 minutes (also in a blocked design). This recent success, with only brief scanning sessions seems to the the most promising method. Therefore it was the method we used to distinguish the two subfields.

Wide Field Retinotopy

Since this experiment began, an additional method (Yan & Wu, 2010; Yan et al., 2011) has been suggested for distinguishing the two regions. By using a very nearby screen, and contact lenses to retain correct focus, the researchers were able to present very wide fields of stimuli, 120° across. They found MT responded more reliably to the most eccentric stimuli. This paradigm did not (in itself) appear to be particularly suited to localisation. The method may be useful in extending the distance from the centre-line in the ipsilateral presentation method, ensuring MT is not activated.

A.2 Localisation Results

A.2.1 hMT+

Most clusters were found at the locations described by previous papers (see methods). Occasionally exceptions were found. For example, comparing the clusters and anatomy of participant LCA24 to JME30, we can see that the relevant sulcus is considerably more ventral (figure A.1). Table A.1 records the cluster centre talairach coordinates for hMT+.

A.2.2 A1

Table A.2 lists the coordinates of A1 across participants. Incidentally, it was noticed that the left and right cluster sizes appeared to be different. Ignoring subjects KUK23 and SMW29b (who had different thresholds applied to the opposite hemispheres). We tested the hypothesis that the left A1 clusters would be significantly larger, as previous research suggests both anatomically (Rademacher et al., 2001) and functionally in the scale of the BOLD response (Devlin et al., 2003). This result was indeed found (mean size, right: 91 voxels, left: 142 voxels; p = 0.032. z = -1.85. ranksum = 80. one-sided wilcoxon rank-sum test).

Participant	Hemisphere	p threshold	Cluster Centre			Voxels	Notes
			x	y	z		
LCA24	Right	0.01 (uc)	40	-70	-8	?	lcg
	Left	0.005 (uc)	-44	-65	3	153	
JME30	Right	0.05	48	-68	9	?	[unused]
	Left	0.05	-46	-63	11	?	Used sphere (r=6mm)
CMR15	Right	0.05	54	-64	7	164	lcg
	Left	0.05	-46	-68	5	?	Used sphere (r=10mm)
SLN11	Skipped						[unused]
ECN13	Right	0.05	54	-63	8	126	lcg
	Left	0.05	-43	-72	6	188	
KHS22	Right	0.05	45	-65	2	86	lcg(Rev)
	Left	0.05	-45	-63	6	120	
MHD12	Right	0.05	43	-44	-38	140	lcg(Rev)
	Left	0.05	-49	-43	-37	169	
EKW30	Right	0.05	52	-60	9	96	lcg
	Left	0.05	-49	-73	12	98	
SMW29	Right	0.05	45	-65	3	206	lcg
	Left	0.05	-45	-74	6	278	
KUE24	Right	1e-9	56	-58	10	409	[unused]
	Left	1e-9	-44	-71	13	184	
KUK23	Right	0.05	46	-77	2	121	hcg
	Left	1e-8	-46	-77	-3	78	
SMW29b	Right	0.05	47	-70	-1	170	hcg
	Left	0.05	-44	-67	4	170	
FLS13	Right	0.01 (uc)	38	-71	2	37	[unused]
	Left	0.01 (uc)	-50	-72	-4	94	
HBA02	Right	0.05	43	-72	3	99	hcg
	Left	0.05	-47	-76	4	154	
DGN29	Right	0.05	56	-64	-1	50	hcg
	Left	0.002 (uc)	-39	73	3	105	
MPA01	Right	0.05	48	-64	6	?	hcg
	Left	0.05	-45	-75	4	197	

Table A.1: hMT+ cluster coordinates. hcg = high contrast group, lcg = low contrast group. [unused] = subject was unused in full analyses. (uc) = uncorrected for multiple comparisons. ? = number of voxels not recorded. Rev = participant had auditory left/right signals reversed.

Participant	Hemisphere	p threshold	Cluster Centre			Voxels
			x	y	z	
CMR15	Right	0.05	63	-47	14	50
	Left	0.05	-63	-39	20	93
ECN13	Right	0.05	55	-12	0	330
	Left	0.05	-47	-27	8	401
KHS22	Right	0.05	57	-26	10	89
	Left	0.05	-51	-24	6	72
MHD12*	Right	0.005uc	48	-22	8	37
	Left	0.005uc	-51	-45	10	130
EKW30*	Right	0.05	57	-32	20	69
	Left	0.05	-47	-31	6	142
SMW29	Right	0.05	53	-23	9	144
	Left	0.05	-46	-24	4	39
KUK23	Right	0.0001uc	62	-15	11	72
	Left	0.05	-48	-26	9	117
SMW29b	Right	0.05	59	-15	15	150
	Left	0.001uc	-53	-30	27	80
HBA02	Right	0.05	60	-22	6	42
	Left	0.05	-51	-30	12	90
DGN29	Right	0.05	63	-12	10	48
	Left	0.05	-59	-28	9	209
MPA01	Right	0.05	61	-16	9	78
	Left	0.05	-60	-21	12	116
LCA24	Right	0.005uc	64	-35	15	27
	Left	0.005uc	-51	-44	10	130

Table A.2: A1 cluster coordinates. MHD12 and EKW30 (*) only used first unimodal session so others could be used in the full analysis. uc = uncorrected for multiple comparisons.

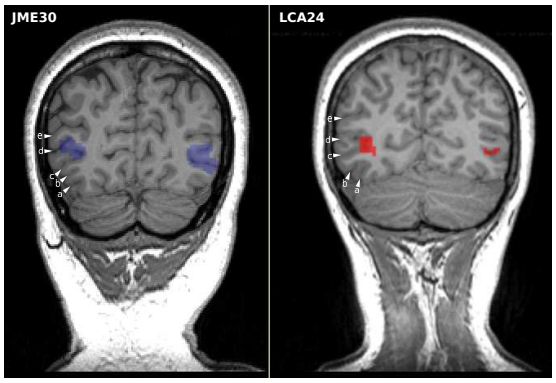


Figure A.1: Participant LCA24 had considerably more ventral cluster of activity for hMT+ contrast. Inspection of sulci suggests that all the sulci were more ventral, consistent with the cluster's location.

A.2.3 VIP

See table A.3 for VIP cluster coordinates.

A.2.4 pV6

pV6 was localised using the same functional localiser as used for VIP (coherent egomotion compatible dot motion vs random dot motion). Table A.4 describes the location of the clusters for each participant. Figure A.2 illustrates the region's location.

A.2.5 pCSv

Participant	Hemisphere	p threshold	Cluster Centre			Voxels	Notes
			x	y	z		
JME30	Right	0.00001uc	28	-52	59	101	
	Left	0.0001uc	-27	-49	55	127	
CMR15	Right	0.05	26	-64	47	?	used sphere to assist constraint (r=6mm)
	Left	0.05	-23	-49	12	?	
ECN13	Right	0.001uc	21	-56	60	120	
	Left	0.001uc	-22	-61	61	121	
KHS22	Right	0.001uc	30	-43	53	56	
	Left	0.001uc	-30	-51	61	50	
MHD12	Right	0.01uc	27	54	58	21	
	Left	0.02uc	-36	-48	59	44	
EKW30	Right	0.001uc	47	-44	-45	26	
	Left	0.001uc	-49	-73	12	98	
KUK23	Right	0.05uc	25	-52	54	39	
	Left	0.001uc	-31	-46	60	34	
SMW29b	Right	0.02uc	18	-55	52	32	
	Left	0.05uc	-25	-52	55	44	Supplemented VIP contrast with hMT+ contrast to assist in localising.
HBA02	Right	0.0001uc	34	-56	59	35	
	Left	0.001uc	-35	-60	58	48	
DGN29	Right	0.05uc	23	-51	53	37	
	Left	0.045uc	-26	-44	47	34	Threshold set to 0.045 to avoid ‘bridge’ to other clusters.
MPA01	Right	0.02uc	-33	-57	52	35	
	Left	0.02uc	24	-52	51	45	

Table A.3: VIP cluster coordinates. uc = uncorrected for multiple comparisons.

Participant	Hemisphere	p threshold	Cluster Centre			Voxels	Notes
			x	y	z		
SMW29	Right	0.05uc	21	-71	23	47	
	Left	0.005uc	-25	-72	36	67	
EKW30	Right	0.005uc	20	-70	59	37	
	Left	0.001uc	-25	-72	36	67	
LCA24	Right	0.01uc	14	-77	35	33	
	Left	0.05	-23	-80	27	40	
CMR15	Right	0.05FWE	22	-79	42	24	
	Left	0.05FWE	-24	-83	37	53	
ECN13	Right	0.01uc	9	-69	26	15	
	Left	0.001uc	-11	-73	27	26	
KHS22	Right	5×10^{-5}uc	22	-73	34	70	
	Left	5×10^{-5}uc	-22	-87	25	78	
MHD12	Right	0.001uc	23	-73	26	22	
	Left	0.001uc	-19	-73	12	27	
KUK23	Right	0.001uc	11	-77	39	32	
	Left	0.001uc	-24	-83	20	21	
DGN29	Right	0.01uc	22	-79	27	32	
	Left	0.01uc	-14	-76	36	8	
MPA01	Right	0.05FWE	28	-73	20	?	
	Left	0.05FWE	-29	-82	20	48	
HBA02	Right	0.05FWE	26	-73	27	68	
	Left	0.002uc	-23	-84	30	20	

Table A.4: V6 cluster coordinates. uc = uncorrected for multiple comparisons, FWE = corrected for familywise error rate.

Participant	Hemisphere	p threshold	Cluster Centre			Voxels	Notes
			x	y	z		
SMW29	Right	0.05uc	17	-13	52	33	
	Left	0.05uc	-8	-17	54	2	
EKW30	Right	0.005uc	15	-19	41	?	
	Left	0.005uc	-12	-24	35	?	
LCA24	Right	-	-	-	-	-	Not Found
	Left	-	-	-	-	-	Not Found
CMR15	Right	0.005uc	8	-39	49	74	
	Left	0.05uc	-11	-24	42	27	
ECN13	Right	0.05uc	15	-25	45	16	
	Left	0.05uc	-13	-21	40	5	
KHS22	Right	0.05uc	3	-19	43	?	
	Left	0.05uc	-9	-19	48	?	
MHD12	Right	0.05uc	11	-23	44	14	
	Left	0.05uc	-14	-29	43	38	
KUK23	Right	0.05uc	12	-36	45	11	
	Left	0.05uc	-13	-28	43	?	
DGN29	Right	0.05uc	9	-18	39	53	
	Left	0.05uc	-8	-15	38	8	
MPA01	Right	0.01uc	18	-33	38	23	
	Left	0.01uc	-13	-29	47	13	
HBA02	Right	-	-	-	-	-	Not Found
	Left	0.01uc	-13	-29	47	13	

Table A.5: pCSv cluster coordinates. uc = uncorrected for multiple comparisons, FWE = corrected for familywise error rate.

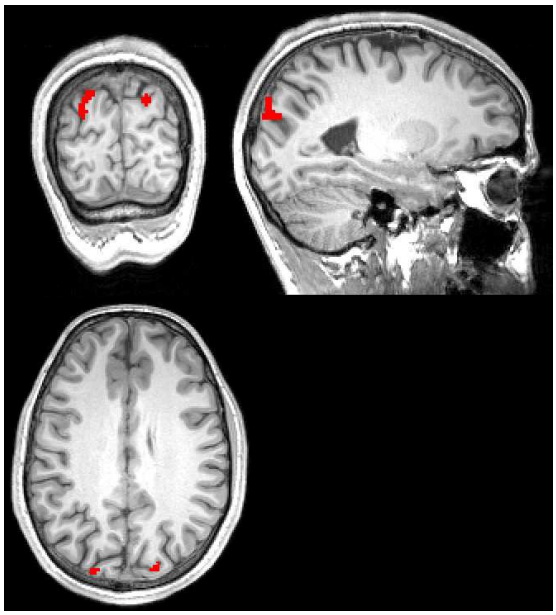


Figure A.2: Location of pV6 as defined by the functional localiser (Participant CMR15, using a $p < 0.05$ FWE mask).

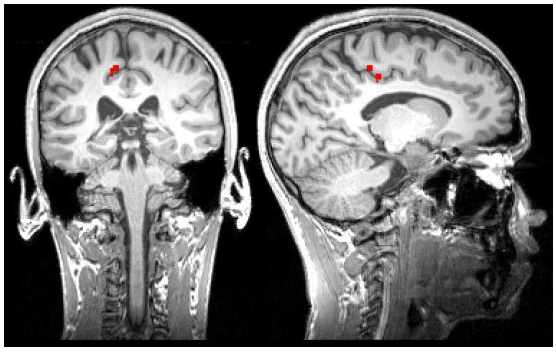


Figure A.3: Location of pCSv as defined by the functional localiser (Participant CMR15, using a $p < 0.005$ uncorrected mask).

Appendix B

Head-direction fMRI study: Localising

Table B.1 summarises the statistical significance thresholds used to decode the three regions.

B.1 The Retrosplenial Cortex and the Posterior Cingulate

Figure B.1 illustrates the localisation of the retrosplenial cortex (functionally) and the posterior cingulate (anatomically). It shows a clear discrepancy. It is likely, I believe, that the functional localiser is responding to activity in other areas.

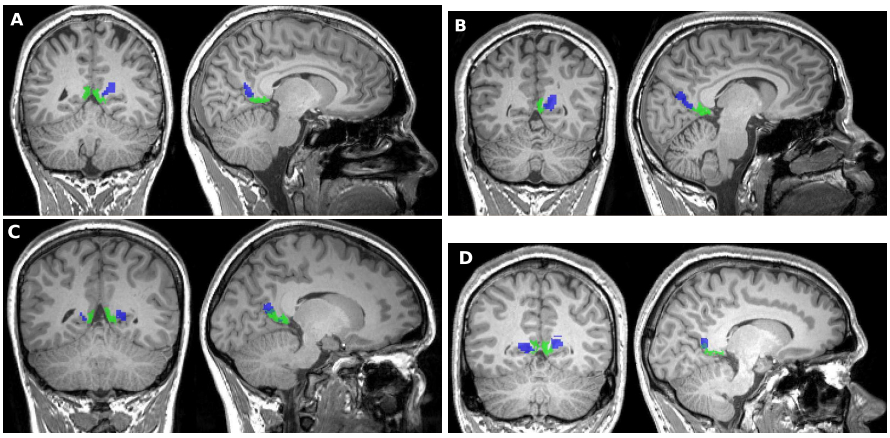


Figure B.1: Example from four participants (A. YBR17, B. TMR04, C. SSA22, D. MLM23) of the functionally defined retrosplenial cortex (blue) and anatomically defined posterior cingulate (green).

Participant	Region		
	FFA	PPA	RSP
AWA13	0.001uc	0.05+sphere	0.05+sphere
LWH13	-	-	-
TMR04	0.005uc	0.05	0.001uc+sphere
SSA22	-	0.001+sphere	0.005
ECN13	0.05	0.05uc	0.05uc
MLM23	0.05uc	0.05	0.05
JHH19	0.05	0.05uc	0.05uc
CPM27	0.05	0.05	0.001uc
YBR17	0.05	0.01+sphere	0.02
LMN22	0.05	0.01+sphere	0.001uc+pshere

Table B.1: For each region localised a statistical threshold was chosen to isolate the relevant cluster of voxels. Here this is represented as the p value used. In most cases this was corrected using FWE multiple comparisons correction. uc indicates the cases which were uncorrected. '+sphere' indicates if a sphere was also used to restrict the ROI.

Appendix C

Decoding and Classification

C.1 Classifiers

Nine classifiers were developed or investigated for use in the analysis of this experiment (Table C.1). Different classifiers have different benefits depending on the type of data being classified. I'll summarise the types used here, starting with the most simple: Logistic Regression, Ridge Regression and Nearest Neighbour.

Logistic Regression has the advantage of being the most simple classifier and the importance of each voxel is represented in the weight each voxel is given. The simplicity of the classification (with flat decision boundaries) reduces the risk of overfitting. However, with a large number of input dimensions and only a few trials, the problem is often underconstrained, causing the $A^\top A$ matrix to become singular (where A is the data matrix), which means there are multiple decision surfaces with the same cost which makes it difficult for the algorithm to select a weight vector (as the inverse becomes difficult to find).

A modification known as ridge regression (a type of Tikhonov-Miller regularization) adds an additional constraint to the problem to allow the classifier to converge on a single solution - that the length of the L^2-norm of the parameters, x, must not exceed a particular value. This allows underconstrained problems to still be given unique solutions by assuming that smaller weights are *a priori* more likely. This does not completely solve the problem of underconstrained data, and in some ways is a fix for the more general problem of the curse of dimensionality - that there are too many dimensions for the number of data points. Additional, more fMRI specific solutions to this problem are mentioned in section C.1.2.

Another very basic classifier developed for this study was the k-nearest neighbour classifier. In this classifier the test point is classified based on trial type of the nearest k samples in the training data. This classifier was included as there was concern that the analysis was suffering from overfitting in some cases. However, it was, anecdotally, not found to be very successful compared to SVM.

The classification toolbox (MVPA) was provided with a support vector machine classifier, however this only could classify between two classes. Chih-Chung Chang and Chih-Jen Lin have created a

Classifier	Description and observations
Logistic Regression	Simple, not usually very successful. Poor when there is little data compared to dimensions.
Ridge Regression	Regularised version of Logistic Regression, allows classification to continue in under constrained conditions.
Nearest Neighbour	Very simple (usually quite poor) classifier, which classifies based on euclidean distances to the set of nearby known data points.
(linear) Support Vector Machine	A very widely used, robust, classifier. Generally not too susceptible to over fitting.
LARS/LASSO	LASSO regularisation means most inputs will not be included in the classification. Useful if most of the voxels are irrelevant. LARS has similar results to LASSO.
Partial Least Squares	Method used by Rodríguez (2012), particularly suited to high-dimensional, correlated data with few samples.
Gaussian Naive Bayes	A fairly simple classifier. Trains a generative model with Gaussian features.
Restricted Boltzmann Machines	Generative classifier, trains a RBM for each class.

Table C.1: The classifiers used during the analysis.

simple to use and open SVM classifier (Chang & Lin, 2011)[1]. By default we used the linear support vector machine. Over the course of my PhD, I found the lSVM to be the most reliable and successful classifier, in general, although other classifiers did do better on some datasets. Two criticisms of SVM classification are that the parameters are less interpretable, and the classifier is more complex than simple linear regression, with a very slightly greater risk of overfitting.

In some cases it might be expected that only a small number of voxels are providing a useful signal, with most restricted to being purely noise. For example if the ROI is placed with uncertainty, it may include cortex outside the intended region. Similarly a large region such as the hippocampus may only contain a small portion with place-related activity. One approach involves a similar method to the ridge regression algorithm; but rather than constrain the L^2-norm, constrain the L^1-norm. This has the effect of reducing the number of non-zero parameters and is known as LASSO regularisation. Figure C.1 illustrates this effect. It means that, of hundreds of voxels, only a few dozen might be included in the classification. This may fit with our expectations about the data, and second provide a useful way of visualising which voxels are most relevant. It it makes most sense to apply LASSO regularisation to those data in which only a few voxels are relevant. For example, almost all voxels will have useful information when decoding whether or not there is motion (from hMT+), which means a different classifier might be better suited. Efron et al. (2004) reported that LASSO (and another method called Stagewise linear regression) are both connected to their algorithm, Least Angle Regression (LARS).

[1](from the README) 'The MATLAB interface was initially written by Jun-Cheng Chen, Kuan-Jen Peng, Chih-Yuan Yang and Chih-Huai Cheng from Department of Computer Science, National Taiwan University. The current version was prepared by Rong-En Fan and Ting-Fan Wu'. A wrapper was created by Ryan Mruczek to allow it to interface with the MVPA toolbox

LARS and LASSO were both interfaced with the MVPA classifier by writing wrapper matlab functions for Karl Skoglund's MATLAB script (based on Efron et al., 2004). An additional phase of k-fold optimisation is performed within the training data, to find the optimum number of voxels.

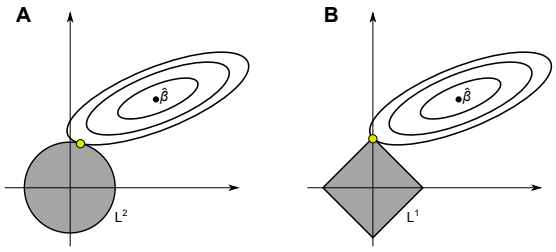


Figure C.1: The elliptical contours indicate equal cost points for the two parameters (along the x and y axes). To regularise, the solution is constrained to be within the filled region. A. Ridge-regression (L^2 regularisation): The constraint is that the sum of the square of the parameters is less than a particular value. This equates to the distance which means the regularised region is a hyper-sphere. B. LASSO L^1-regularisation: The constraint is that the sum of the parameters is less than a particular value. This means that the region becomes a hyper-cube. The upshot is that the vertices or edges of the hypercube are more likely to intersect the contour of the cost function before the edges or surfaces. At the vertices or edges, one or more of the parameters is zero.

Rodriguez (2010) used a Partial Least Squares (PLS) classifier and argued it was particularly relevant for decoding place from the hippocampus due to its particular utility when there are more variables than observations (a common problem in fMRI, especially when the ROI contains potentially thousands of voxels, such as the hippocampus). It also handles the situation well when some of the variables are correlated (which is also a feature of fMRI voxels). It was logical to use this classifier for our data. The MATLAB stats toolbox contained a PLS implementation. I wrote a wrapper for the MVPA toolbox to use the PLS classifier, again using cross-validation to select the optimum parameter values.

The MVPA toolkit supplied a Gaussian Naive Bayes classifier, which was tested during development. An additional classifier, developed for this study, used a series of restricted boltzmann machines, each trained on a different class. Classification of the test data was performed by selecting the machine with the lowest free energy. The tool was found to classify somewhat worse than SVM, this may have been due to a bias towards some classes more than others. It was also unoptimised and therefore very slow. It was based on various work by Geoffrey E. Hinton (e.g. Schmah et al., 2009).

C.1.1 Image preprocessing

Rather than simply use the fMRI data, the images can first be processed using the SPM toolbox. For example, beta images from a GLM could be used instead, with each beta image the result of a regressor representing one of the conditions in one block. This can potentially improve the signal by taking into account the haemodynamic response function, other regressors (explaining parameters about the

experiment or head-motion) and allows the signal from multiple trials to be combined. Taking into account the haemodynamic response is particularly important in an event driven design like the one in this study. Mumford et al. (2012) found that using a general linear model with each trial as a regressor improved the classification accuracy in their example data.

An alternative, intermediate option involves creating a regressor containing motion and other artefacts or distractor regressors. The residual ('error') is then used for classifying the data. This in effect 'regresses out' linear components of head-motion etc.

C.1.2 Additional Methods

Beyond the basic classifier several other techniques were employed in an attempt to detect the head-direction and place signal. To fully investigate these options, visualise the results and work on the analysis in a systematic way, a MATLAB software tool was developed which allows classification computations to be recorded. See the C.1.3 section below for more information about this tool.

Searchlight

As with the Hassabis et al. (2009) study, rather than use the whole ROI for decoding, a small spherical region within the ROI is selected and the cross-validation performed on the signal from this small neighbourhood. This is repeated separately for spheres centred on each voxel. Multiple comparison correction is then performed on the results of the classifications.

Searchlight analysis has several advantages. First it is a compromise between covering a large volume which might be necessary if it is unclear where the signal of interest is, and restricting the number of inputs (or dimensions) to the classifier. In particular if the number of dimensions is more than the number of trials then the classification problem becomes very difficult. This general problem is often referred to as the 'curse of dimensionality'; the data points become sparse and difficult to group or classify. By reducing the input to a set of 100 or so voxels, the classifier may find it easier to extract any signal than if the same data was hidden in larger population of noisy voxels.

In this study we used the conservative bonferroni correction. However, it is worth noting that there is considerable correlation between neighbouring spheres (due to them sharing many voxels, and the underlying correlation between voxels). Such correlation between sphere results can be enforced by spatially smoothing the results and using a random effects analysis, as in Morgan et al. (2011). Alternatively one can ignore multiple comparisons correction and instead compare the frequency of significant spheres between two regions, as in Hassabis et al. (2009, supplemental data).

Remove Averaging

Another way to reduce the problem of dimensionality is to increase the number of samples. The data in classification studies is usually averaged across trials in the same block. However this is not necessary and, although increasing the noise within each data point, the volumes can be used separately. It is more

important that an empirically derived null distribution is calculated for this analysis, as the correlation between trials will be greater (it is also important that the cross-validation groups do not divide trials of the same block across groups, as these will be heavily correlated).

t-mask

To further reduce the number of voxels included in the classification, one of the blocks used for training could be removed and instead used to provide a t-statistic for each voxel (by performing a simple univariate general linear model against the regressors). The results of this test are then thresholded and the voxels which achieved an uncorrected significance are then used in the standard cross-validation. In this way a second tier of cross-validation takes place, using the t-mask to select the most salient voxels for classification, to aid the classifier when analysing large volumes.

Smoothing

One way to reduce noise is to average across neighbouring voxels by spatially smoothing the data. Both Kamitani & Sawahata (2010) and Morgan et al. (2011) smoothed the data prior to classification. This, on the face of it, appears to mix together different voxels and thus weaken classification. However, Kamitani & Sawahata (2010) investigates this in considerable detail and conclude that the harm caused by the smoothing is equalled by the benefits it offers in noise-reduction.

'Stacking' volumes

The start times of the blocks are delayed by several TRs, to model the delay caused by the sluggish haemodynamic response. The delay length varies between studies. By default we use 2 TRs (4 seconds), but 1,3 or 4 may function better. Which of the volumes during the trial period are most informative is also not known. A modification made to the classifier [2] involves taking the individual volumes from a trial and combining them into one 'super-volume', consisting of four or five volumes of data from that whole trial. If, for example, the later responses are most informative, then the classifier will be able to use the data principally from those later trials and ignore the earlier ones. This method allows us to leverage the power of the classifier to determine which time points are most relevant. Ideally the stimulus would begin at a particular offset time relative to the acquisition periods, so each volume is synchronised with the relevant volume from other trials.

C.1.3 The Decoding Tool

To help organise the results of classification and decoding, I created a simple MATLAB tool to assist in visualising the results. Figure C.2 illustrates the tool with the key features highlighted. From this console one can start new analyses, review previous classification results both at group, subject and

[2] suggested by Amos Storkey, who kindly spent a considerable period of time looking at the experimental results and our analysis.

voxel level. Figure C.3 shows how, when one clicks on a searchlight analysis the tool displays the distribution of searchlight results. Clicking on the 'View MRI' button opens MRIcron with the ROI highlighted, and within the ROI the voxels which reach 5% (corrected) significance (see figure C.4).

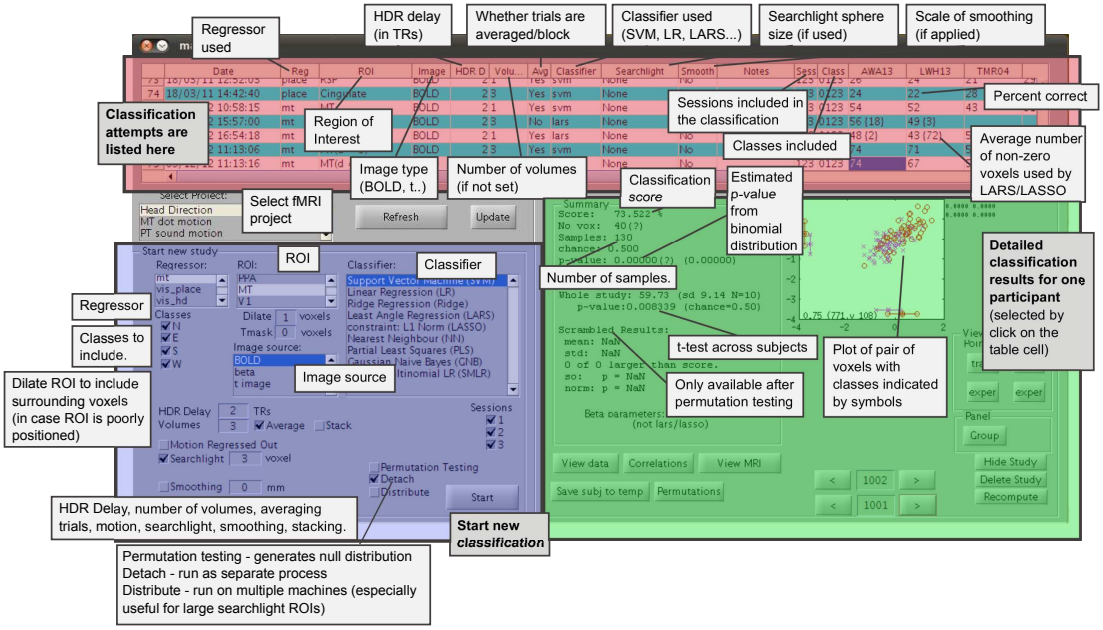


Figure C.2: The tool used for organising the decoding and analysis for all the classification experiments. The main window is split into three parts. At the top is a table of all the classification attempts made, the configuration of each and the results for each participant. On the left is a console for starting new classification runs. Here one can configure almost all the key features of interest, including: the regressor, region, classifier, t-mask, image-type, classes and session to use, HDR delay, searchlight size, smoothing, etc. On the right are details about a particular subject's results. Here one can plot individual pairs of voxel data, view MRI imagery (for LARS/LASSO and searchlight results), and see summaries of statistical tests.

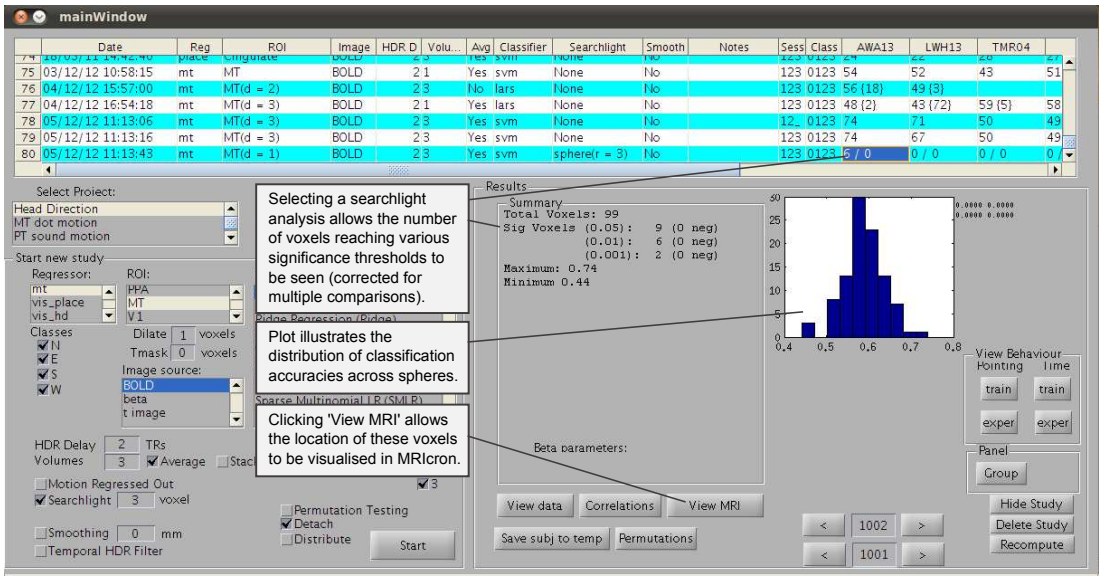


Figure C.3: To investigate a searchlight analysis the tool plots a histogram showing the distribution of accuracy across spheres. In this case 99 spheres were classified, each with a radius of 3 voxels. Nine of these had significance (once multiple comparisons correction is included) below the 5% level.

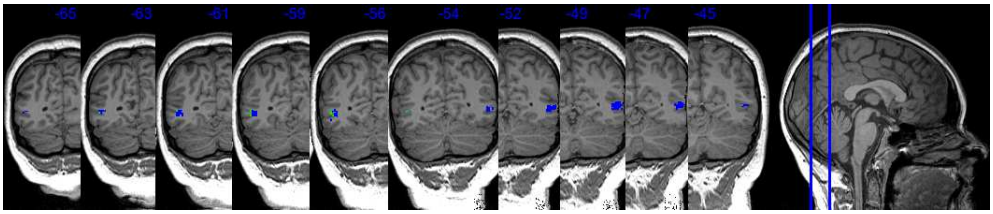


Figure C.4: By clicking on the 'View MRI' button, users can view where the significant searchlight sphere centres are (green) voxels. This method also allows the voxels used by the LARS/LASSO classifier to be viewed (LARS/LASSO set many or most of the coefficients to zero, and therefore only use a small subset of voxels when classifying). Image shows coronal sections for subject AWA13 (as illustrated in figure C.3).

Appendix D

Simple Planar Motion: Localising

Although smaller than V1, because the hMT+ mask was functionally defined using a smoothed map, it did not follow the cortical surface and therefore had more voxels in than one would expect, compared to the V1 mask, which was anatomically defined, and neatly traced the cortex. The hMT+ mask had, on average, 545 voxels (bilateral), which is equivalent to $4,000mm^3$. The V1 region had, on average, 603 voxels, which is equivalent to $4,434mm^3$.

Regions hMT+ was first localised using the functional fMRI sequence data, and then 160 voxels were selected from this ROI by using the t-mask of that initial localiser. There was concern expressed about the size of the region compared to the 160 voxels selected: The voxels were $7.35mm^3$ each, which meant that 80 voxels would consist of $588mm^3$ or a cube approximately $8mm$ on each side. To illustrate how these voxels are selected from the larger mask, figure D.1 illustrates for a single subject the initial mask (blue) and the 160 voxels selected from it (green). To give an idea of the variation in the hMT+ ROIs, its mean, as mentioned, was of 545 voxels, the standard deviation in the number of voxels in each ROI is 227, with a range of between 216 and 941, median 515 and inter-quartile range 403. The variation is due largely to the choice of threshold for the t-mask, which needs to be varied between participants to take into account the differences in the signal-to-noise ratio. In retrospect it might have been wise to attempt to make the mask volumes more similar across subjects. For comparison the standard deviation in the size of V1 was only 81 voxels (mean, 603), reflecting the similarity in anatomically defined ROIs.

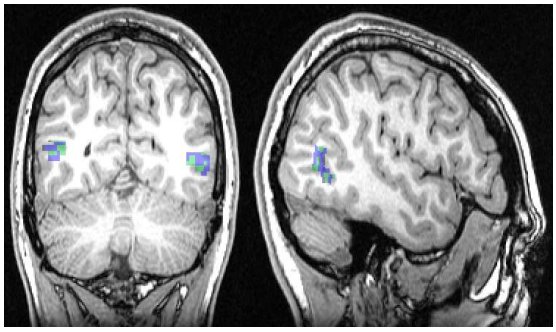


Figure D.1: Participant s05 showing the initial large hMT+ mask (in blue) and the smaller selection of 160 voxels used by the classifier (in green).

Bibliography

Adams, W., Gray, K., Garner, M., & Graf, E. (2010). High-level face adaptation without awareness. *Psychological Science*, *21*(2), 205–210.

Aguirre, G. K., Mattar, M. G., & Magis-Weinberg, L. (2011). de Bruijn cycles for neural decoding. *Neuroimage*, *56*(3), 1293–1300.

Ahissar, M., Ahissar, E., Bergman, H., & Vaadia, E. (1992). Encoding of sound-source location and movement: activity of single neurons and interactions between adjacent neurons in the monkey auditory cortex. *Journal of neurophysiology*, *67*(1), 203–215.

Alais, D., & Burr, D. (2004). The Ventriloquist Effect Results from Near-Optimal Bimodal Integration. *Current Biology*, *14*(3), 257–262.

Albright, T. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*, *52*, 1106–1130.

Alink, A., Euler, F., Kriegeskorte, N., Singer, W., & Kohler, A. (2011). Auditory motion direction encoding in auditory cortex and high-level visual cortex. *Human brain mapping*, *33*, 969–978.

Alink, A., Singer, W., & Muckli, L. (2008). Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *The Journal of Neuroscience*, *28*(11), 2690–2697.

Allen, P. G., & Kolers, P. A. (1981). Sensory Specificity of Apparent Motion. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(6), 1318–1326.

Allman, J., & Kass, J. (1975). The dorsomedial cortical visual area: A third tier area in the occipital lobe of the owl monkey (¡i¿aotus trivirgatus¡/i¿). *Brain research*, *100*(3), 473–487.

Andersen, P., Morris, R., Amaral, D., & O'Keefe, J. (2007). *The Hippocampus Book (1st ed.)*. Oxford University Press US.

Angelaki, D., Gu, Y., & DeAngelis, G. (2009). Multisensory integration: psychophysics, neurophysiology and computation. *Current opinion in neurobiology*, *19*(4), 452.

Ballard, C., Mohan, R., Bannister, C., Handy, S., & Patel, A. (1991). Wandering in dementia sufferers. *International Journal of Geriatric Psychiatry*, *6*(8), 611–614.

Barlow, H., & Hill, R. (1963). Evidence for a physiological explanation of the waterfall phenomenon and figural after-effects. *Nature*.

Bassett, J. P., & Taube, J. S. (2001). Neural correlates for angular head velocity in the rat dorsal tegmental nucleus. *The Journal of Neuroscience*, *21*(15), 5740–5751.

Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A.*, *20*(7), 1391–1397.

Baumann, O., & Greenlee, M. (2007). Neural correlates of coherent audiovisual motion perception. *Cerebral Cortex*, *17*, 1433–1443.

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

Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Current opinion in neurobiology*, *15*, 145–153.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*(1), 149–159.

Bedny, M., Konkle, T., Pelphrey, K., Saxe, R., & Pascual-Leone, A. (2010). Sensitive period for a multimodal response in human visual motion area MT/MST. *Current Biology*, *20*, 1900–1906.

Beer, A., & Röder, B. (2004). Unimodal and crossmodal effects of endogenous attention to visual and auditory motion. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 230–240.

Benevento, L., Fallon, J., Davis, B., & Rezak, M. (1977). Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Experimental neurology*, *57*(3), 849–872.

Bett, D., Stevenson, C., Shires, K., Smith, M., Stephen, J., Dudchenko, P., & Wood, E. (2013). The postsubiculum and spatial learning: the role of postsubicular synaptic activity and synaptic plasticity in hippocampal place cell-, object- and object-location memory. *Journal of Neuroscience*, *in press*.

Bolan, P. J., Yacoub, E., Garwood, M., Ugurbil, K., & Harel, N. (2006). In vivo micro-mri of intracortical neurovasculature. *Neuroimage*, *32*(1), 62–69.

Bonnici, H., Chadwick, M., Kumaran, D., Hassabis, D., Weiskopf, N., & Maguire, E. (2012). Multi-voxel pattern analysis in human hippocampal subfields. *Frontiers in Human Neuroscience*, *6*, 290.

Bonnici, H., Kumaran, D., Chadwick, M., Weiskopf, N., Hassabis, D., & Maguire, E. (2011). Decoding representations of scenes in the medial temporal lobes. *Hippocampus*, *22*(5), 1143–1153.

Boussaoud, D., Ungerleider, L., & Desimone, R. (1990). Pathways for motion analysis: cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque. *The Journal of comparative neurology*, *296*, 462–495.

Boynton, G., & Finney, E. (2003). Orientation-specific adaptation in human visual cortex. *The Journal of neuroscience*, *23*(25), 8781–8787.

Bremmer, F. (2005). Navigation in space–the role of the macaque ventral intraparietal area. *The Journal of physiology*, *566*(1), 29–35.

Bremmer, F., Duhamel, J., Hamed, S. B., & Graf, W. (2002a). Heading encoding in the macaque ventral intraparietal area (VIP). *European Journal of Neuroscience*, *16*, 1554–1568.

Bremmer, F., Klam, F., Duhamel, J., Hamed, S. B., & Graf, W. (2002b). Visual-vestibular interactive responses in the macaque ventral intraparietal area (VIP). *European Journal of Neuroscience*, *16*, 1569–1586.

Bremmer, F., Kubischik, M., Pekel, M., Hoffmann, K., & Lappe, M. (2010). Visual selectivity for heading in monkey area MST. *Experimental brain research*, *200*, 51–60.

Bremmer, F., Schlack, A., Shah, N., Zafiris, O., Kubischik, M., Hoffmann, K., Zilles, K., & Fink, G. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron*, *29*(1), 287–296.

Britten, K. H. (1998). Clustering of response selectivity in the medial superior temporal area of extrastriate cortex in the macaque monkey. *Visual neuroscience*, *15*(3), 553–558.

Britten, K. H. (2008). Mechanisms of self-motion perception. *Annual Review of Neuroscience*, *31*, 389–410.

Burr, D., & Thompson, P. (2011). Motion psychophysics: 1985–2010. *Vision research*, *51*(13), 1431–1456.

Buxton, R. B., Uludağ, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage*, *23*, S220–S233.

Byrne, P., Becker, S., & Burgess, N. (2007). Remembering the past and imagining the future: a neural model of spatial memory and imagery. *Psychological Review*, *114*(2), 340.

Calabro, F., Soto-Faraco, S., & Vaina, L. (2011). Acoustic facilitation of object movement detection during self-motion. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1719), 2840–2847.

Calton, J., et al. (2003). Hippocampal place cell instability after lesions of the head direction cell network. *The Journal of Neuroscience*, *23*(30), 9719–9731.

Cardin, V., & Smith, A. (2010). Sensitivity of human visual and vestibular cortical regions to egomotion-compatible visual stimulation. *Cerebral Cortex*, *20*, 1964–1973.

Casanova, M., Buxhoeveden, D., Switala, A., & Roy, E. (2002). Asperger's syndrome and cortical neuropathology. *Journal of child neurology*, *17*(2), 142–145.

Chang, C. C., & Lin, C. J. (2011). LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology. Software available at http://www.csie.ntu.edu.tw/c̃jlin/libsvm*, *2*(3), 27.

Chapman, B., & Bonhoeffer, T. (1998). Overrepresentation of horizontal and vertical orientation preferences in developing ferret area 17. *Proceedings of the National Academy of Sciences*, *95*(5), 2609–2614.

Chen, L., Lin, L., Green, E., Barnes, C., & McNaughton, B. (1994). Head-direction cells in the rat posterior cortex. *Experimental Brain Research*, *101*(1), 8–23.

Cheng, K., Huttenlocher, J., Shettleworth, S. J., & Rieser, J. J. (2007). Bayesian integration of dpatial information. *Psychological Bulletin*, *133*(4), 625–637.

Cheng, K., Waggoner, R., & Tanaka, K. (2001). Human ocular dominance columns as revealed by high-field functional magnetic resonance imaging. *Neuron*, *32*(2), 359–374.

Cheung, T. H. C., & Cardinal, R. N. (2005). Hippocampal lesions facilitate instrumental learning with delayed reinforcement but induce impulsive choice in rats. *BMC Neuroscience*, *6*(36).

Choi, H., Zilles, K., Mohlberg, H., Schleicher, A., Fink, G., Armstrong, E., & Amunts, K. (2006). Cytoarchitectonic identification and probabilistic mapping of two distinct areas within the anterior ventral bank of the human intraparietal sulcus. *The Journal of comparative neurology*, *495*(1), 53–69.

Colby, C., Duhamel, J., & Goldberg, M. (1993). Ventral intraparietal area of the macaque: anatomic location and visual response properties. *Journal of neurophysiology*, *69*(3), 902–914.

Constable, R., Henkelman, R., et al. (1991). Contrast, resolution, and detectability in MR imaging. *Journal of computer assisted tomography*, *15*(2), 297.

DeAngelis, G. C., & Angelaki, D. E. (2012). chapter 31: Visual-vestibular integration for self-motion perception. In M. M. Murray, & M. T. Wallace (Eds.) *The Neural Bases of Multisensory Processes*. CRC Press, 1st ed.

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

den Ouden, H., Daunizeau, J., Roiser, J., Friston, K., & Stephan, K. (2010). Striatal prediction error modulates cortical coupling. *The Journal of Neuroscience*, *30*(9), 3210–3219.

Deouell, L. Y., Heller, A. S., Malach, R., D'Esposito, M., & Knight, R. T. (2007). Cerebral responses to change in spatial location of unattended sounds. *Neuron*, (pp. 985–996).

Desimone, R., & Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *The Journal of Comparative Neurology*, *248*, 164–189.

Devlin, J. T., et al. (2003). Functional Asymmetry for Auditory Processing in Human Primary Auditory Cortex. *The Journal of Neuroscience*, *23*, 11516–11522.

Diekmann, V., Jürgens, R., & Becker, W. (2009). Deriving angular displacement from optic flow: a fMRI study. *Experimental brain research*, *195*(1), 101–116.

Doeller, C., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*(7281), 657–661.

Dong, C., Swindale, N., Cynader, M., et al. (1999). A contingent aftereffect in the auditory system. *Nature neuroscience*, *2*, 863–866.

Dong, C., Swindale, N., Zakarauskas, P., Hayward, V., & Cynader, M. (2000). The auditory motion-aftereffect: Its tuning and specificity in the spatial and frequency domains. *Attention, Perception, & Psychophysics*, *62*(5), 1099–1111.

Duffy, C., & Wurtz, R. (1991). Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli. *Journal of Neurophysiology*, *65*(6), 1329–1345.

Duffy, C., & Wurtz, R. (1995). Response of monkey MST neurons to optic flow stimuli with shifted centers of motion. *The Journal of Neuroscience*, *15*(7), 5192–5208.

Duhamel, J., Colby, C., & Goldberg, M. (1998). Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *Journal of neurophysiology*, *79*(1), 126–136.

Dukelow, S. P., DeSouza, J. F., Culham, J. C., van den Berg, A. V., Menon, R. S., & Vilis, T. (2001). Distinguishing subregions of the human MT+ complex using visual fields and pursuit eye movements. *Journal of Neurophysiology*, *86*, 1991–2000.

Eckert, M. A., Kamdar, N. V., Chang, C. E., Beckmann, C. F., Greicius, M. D., & Menon, V. (2008). A cross-modal System linking primary auditory and visual cortices: evidence from intrinsic fMRI connectivity analysis. *Human Brain Mapping*, *29*, 848–857.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of statistics*, *32*(2), 407–499.

Ehrenstein, W. H., & Reinhardt-rutland, A. H. (1996). A cross-modal aftereffect: auditory displacement following adaptation to visual motion. *Perceptual and motor skills*, *82*(1), 23–26.

Eichenbaum, H., Wiener, S. I., Shapiro, M., & Cohen, N. (1989). The organization of spatial coding in the hippocampus: a study of neural ensemble activity. *The Journal of neuroscience*, *9*(8), 2764–2775.

Ekstrom, A., Kahana, M., Caplan, J., Fields, T., Isham, E., Newman, E., & Fried, I. (2003). Cellular networks underlying human spatial navigation. *Nature*, *425*(6954), 184–188.

Engel, S., Glover, G., & Wandell, B. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral cortex*, *7*(2), 181–192.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*(6676), 598–601.

Epstein, R., Parker, W., & Feiler, A. (2007). Where am I now? Distinct roles for parahippocampal and retrosplenial cortices in place recognition. *The Journal of neuroscience*, *27*(23), 6141–6149.

Ernst, M. O. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, *7*(5).

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.

Esiri, M., & Chance, S. (2006). Vulnerability to alzheimer's pathology in neocortex: the roles of plasticity and columnar organization. *Journal of Alzheimer's Disease*, *9*, 79–89.

Felleman, D. J., & Kaas, J. H. (1984). Receptive-field properties of neurons in middle temporal visual area (mt) of owl monkeys. *Journal of Neurophysiology*, *52*(3), 488–513.

Fetsch, C., DeAngelis, G., & Angelaki, D. (2010). Visual–vestibular cue integration for heading perception: applications of optimal cue integration theory. *European Journal of Neuroscience*, *31*, 1721–1729.

Fetsch, C., Turner, A., DeAngelis, G., & Angelaki, D. (2009). Dynamic reweighting of visual and vestibular cues during self-motion perception. *The Journal of Neuroscience*, *29*(49), 15601–15612.

Fetsch, C., Wang, S., Gu, Y., DeAngelis, G., & Angelaki, D. (2007). Spatial reference frames of visual, vestibular, and multimodal heading signals in the dorsal subdivision of the medial superior temporal area. *The Journal of neuroscience*, *27*(3), 700–712.

Fetsch, C. R., Pouget, A., DeAnglelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature Neuroscience*, *15*, 146–154.

Fischer, E., Bülthoff, H. H., Logothetis, N. K., & Bartels, A. (2011). Visual motion responses in the posterior cingulate sulcus: A comparison to V5/MT and MST. *Cerebral Cortex*, *22*, 865–876.

Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–355.

Freeman, J., Brouwer, G., Heeger, D., & Merriam, E. (2011). Orientation decoding depends on maps, not columns. *The Journal of Neuroscience*, *31*(13), 4792–4804.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. (1993). Functional connectivity: the principal component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow & Metabolism*, *13*, 5–14.

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, *19*, 1273–1302.

Geisler, W., et al. (1999). Motion streaks provide a spatial code for motion direction. *Nature*, *400*(6739), 65–68.

Giaschi, D., Zwicker, A., Young, S., & Bjornson, B. (2007). The role of cortical area V5/MT+ in speed-tuned directional anisotropies in global motion perception. *Vision research*, *47*(7), 887–898.

Gibson, J. (1950). *The perception of the visual world*. Oxford, England. Houghton Mifflin.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience*, *14*(7), 926–932.

Glasser, D. M., Tsui, J. M. G., Pack, C. C., & Tadin, D. (2011). Perceptual and neural consequences of rapid motion adaptation. *Proceedings of the National Academy of Sciences*, *108*(45), E1080–E1088.

Golob, E. J., & Taube, J. S. (1997). Head direction cells and episodic spatial information in rats without a hippocampus. *Proceedings of the National Academy of Sciences*, *94*, 7645–7650.

Goodridge, J. P., Dudchenko, P. A., Worboys, K. A., Golob, E. J., & Taube, J. S. (1998). Cue control and head direction cells. *Behavioral Neuroscience*, *112*(4), 749–761.

Goodridge, J. P., & Taube, J. S. (1997). Interaction between the postsubiculum and anterior thalamus in the generation of head direction cell activity. *The Journal of Neuroscience*, *17*(23), 9315–9330.

Grefkes, C., & Fink, G. (2005). Review: The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of anatomy*, *207*(1), 3–17.

Griffiths, T., Rees, A., Witton, C., Cross, P., Shakir, R., & Green, G. (1997). Spatial and temporal auditory processing deficits following right hemisphere infarction. A psychophysical study. *Brain*, *120*, 785–794.

Gros, B., Blake, R., & Hiris, E. (1998). Anisotropies in visual motion perception: a fresh look. *Journal of the Optical Society of America A*, *15*(8), 2003–2011.

Gu, Y., Angelaki, D. E., & DeAnglelis, G. C. (2008). Neural correlates of multisensory cue integration in macaque MSTd. *Nature neuroscience*, *11*, 1201–1210.

Gu, Y., Watkins, P., Angelaki, D., & DeAngelis, G. (2006). Visual and nonvisual contributions to three-dimensional heading selectivity in the medial superior temporal area. *The Journal of neuroscience*, *26*, 73–85.

Hafting, T., Fyhn, M., Molden, S., Moser, M., & Moser, E. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806.

Hansen, B., & Essock, E. (2006). Anisotropic local contrast normalization: The role of stimulus orientation and spatial frequency bandwidths in the oblique and horizontal effect perceptual anisotropies. *Vision research*, *46*(26), 4398–4415.

Hanson, S., Matsuka, T., & Haxby, J. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: Is there a face area? *Neuroimage*, *23*(1), 156–166.

Harvey, C., Collman, F., Dombeck, D., & Tank, D. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature*, *461*(7266), 941–946.

Hassabis, D., Chu, C., Rees, G., Weiskopf, N., Molyneux, P., & Maguire, E. (2009). Decoding neuronal ensembles in the human hippocampus. *Current Biology*, *19*(7), 546–554.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*, 2425–2430.

Hay, L., Bard, C., Fleury, M., & Teasdale, N. (1996). Availability of visual and proprioceptive afferent messages and postural control in elderly adults. *Experimental Brain Research*, *108*(1), 129–139.

Heeger, D., Boynton, G., Demb, J., Seidemann, E., & Newsome, W. (1999). Motion opponency in visual cortex. *The Journal of Neuroscience*, *19*(16), 7162–7174.

Hidaka, S., Teramoto, W., Sugita, Y., Manaka, Y., Sakamoto, S., & Suzuki, Y. (2011). Auditory Motion Information Drives Visual Motion Perception. *PLoS One*, *6*(3), e17499.

Hinds, O., Rajendran, N., Polimeni, J., Augustinack, J., Wiggins, G., Wald, L., Rosas, H., Potthast, A., Schwartz, E., & Fischl, B. (2008). Accurate prediction of V1 location from cortical folds in a surface coordinate system. *Neuroimage*, *39*(4), 1585.

Hoesen, G., & Solodkin, A. (1994). Cellular and systems neuroanatomical changes in Alzheimer's disease. *Annals of the New York Academy of Sciences*, *747*(1), 12–35.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature neuroscience*, *3*(1), 80–84.

Hok, V., Save, E., Lenck-Santini, P., & Poucet, B. (2005). Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(12), 4602–4607.

Hölscher, C., Jacob, W., & Mallot, H. A. (2004). Learned association of allocentric and egocentric information in the hippocampus. *Experimental brain research*, *158*(2), 233–240.

Hubel, D., & Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, *195*(1), 215–243.

Hubel, D., & Wiesel, T. (1974). Sequence regularity and geometry of orientation columns in the monkey striate cortex. *The Journal of comparative neurology*, *158*(3), 267–293.

Huk, A., Dougherty, R., & Heeger, D. (2002). Retinotopy and functional subdivision of human areas MT and MST. *The Journal of Neuroscience*, *22*(16), 7195–7205.

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.

Jehee, J., Brady, D., & Tong, F. (2011). Attention improves encoding of task-relevant features in the human visual cortex. *The Journal of Neuroscience*, *31*(22), 8210–8219.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *Neuroimage*, *62*, 782–790.

Kaas, J., & Hackett, T. (2000). Subdivisions of auditory cortex and processing streams in primates. *Proceedings of the National Academy of Sciences*, *97*(22), 11793–11799.

Kamitani, Y., & Sawahata, Y. (2010). Spatial smoothing hurts localization but not information: pitfalls for brain mappers. *Neuroimage*, *49*(3), 1949–1952.

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature neuroscience*, *8*(5), 679–685.

Kamitani, Y., & Tong, F. (2006). Decoding seen and attended motion directions from activity in the human visual cortex. *Current biology*, *16*(11), 1096–1102.

Kanwisher, N., McDermott, J., & Chun, M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, *17*(11), 4302–4311.

Kaplan, E., Marcus, S., & So, Y. T. (1979). Effects of dark adaptation on spatial and temporal properties of receptive fields in cat lateral geniculate nucleus. *The Journal of physiology*, *294*(1), 561–580.

Karaoguz, C., Weisswange, T., Rodemann, T., Wrede, B., & Rothkopf, C. (2011). Reward-based learning of optimal cue integration in audio and visual depth estimation. In *15th International Conference on Advanced Robotics (ICAR)*, (pp. 389–395). IEEE.

Kasess, C., Stephan, K., Weissenbacher, A., Pezawas, L., Moser, E., & Windischberger, C. (2010). Multi-subject analyses with dynamic causal modeling. *NeuroImage*, *49*(4), 3065–3074.

Kaskan, P., Dillenburger, B., Lu, H., Roe, A., & Kaas, J. (2010). Orientation and direction-of-motion response in the middle temporal visual area (MT) of new world owl monkeys as revealed by intrinsic-signal optical imaging. *Frontiers in neuroanatomy*, *4*(23).

Knierim, J., Kudrimoti, H., & McNaughton, B. (1995). Place cells, head direction cells, and the learning of landmark stability. *The Journal of Neuroscience*, *15*(3), 1648–1659.

Kolster, H., Peeters, R., & Orban, G. (2010). The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *The Journal of Neuroscience*, *30*(29), 9801–9820.

Konen, C. S., & Kastner, S. (2008). Representation of eye movements and stimulus motion in topographically organized areas of human posterior parietal cortex. *Journal of Neuroscience*, *28*(33), 8361–8375.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *Neuroimage*, *37*(4), 1445–1456.

Krekelberg, B., Boynton, G. M., & van Wezel, R. J. A. (2006). Adaptation: from single cells to BOLD signals. *Trends in Neurosciences*, *29*, 250–256.

Lappe, M., Bremmer, F., & Van den Berg, A. (1999). Perception of self-motion from visual flow. *Trends in cognitive sciences*, *3*(9), 329–336.

Larsson, P., Västfjäll, D., & Kleiner, M. (2004). Perception of self-motion and presence in auditory virtual environments. In *Proceedings of 7th annual workshop of presence*, (pp. 252–258).

Lee, D., & Aronson, E. (1974). Visual proprioceptive control of standing in human infants. *Attention, Perception, & Psychophysics*, *15*(3), 529–532.

Leek, M. (2001). Adaptive procedures in psychophysical research. *Attention, Perception, & Psychophysics*, *63*(8), 1279–1292.

Lenck-Santini, P., Muller, R. U., Save, E., & Poucet, B. (2002). Relationships between place cell firing fields and navigational decisions by rats. *The Journal of Neuroscience*, *22*(20), 9035–9047.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*(2), 467–477.

Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A Comparison of Visual and Auditory Motion Processing in Human Cerebral Cortex. *Cerebral Cortex*, *10*, 873–888.

Li, B., Peterson, M., & Freeman, R. (2003). Oblique effect: a neural basis in the visual cortex. *Journal of Neurophysiology*, *90*(1), 204–217.

Logothetis, N. K. (2008). What we can do and what we cannot do with fmri. *Nature*, *453*(7197), 869–878.

Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fmri signal. *Nature*, *412*(6843), 150–157.

Lohmann, G., Erfurth, K., Müller, K., & Turner, R. (2012). Critical comments on dynamic causal modelling. *Neuroimage*, *59*(3), 2322–2329.

Loomis, J. M., Klatzky, R. L., Golledge, R. G., & Philbeck, J. W. (1999). Human navigation by path integration. *Wayfinding behavior: Cognitive mapping and other spatial processes*, (pp. 125–151).

Lu, H. D., Chen, G., Tanigawa, H., & Roe, A. W. (2010). A motion direction map in macaque V2. *Neuron*, *68*(5), 1002–1013.

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*, 1432–1438.

Ma, W. J., & Pouget, A. (2008). Linking neurons to behavior in multisensory perception: a computational review. *Brain Research*, *1242*, 4–12.

Maiworm, M., Bellatoni, M., Spence, C., & Röder, B. (2012). When emotional valence modulates audiovisual integration. *Attention, Perception and Psychophysics*, *74*(6), 1302–1311.

Malonek, D., Tootell, R., Grinvald, A., Malonek, D., Tootell, R., & Grinvald, A. (1994). Optical imaging reveals the functional architecture of neurons processing shape and motion in owl monkey area MT. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *258*(1352), 109–119.

Man, K., Kaplan, J. T., Damasio, A., & Meyer, K. (2012). Sight and sound converge to form modality-invariant representations in temporoparietal cortex. *The Journal of Neuroscience*, *32*(47), 16629–16636.

Mateeff, S., Hohnsbein, J., & Noack, T. (1985). Dynamic visual capture: apparent auditory motion induced by a moving visual target. *Perception*, *14*, 721–727.

Maunsell, J. H., & van Essen, D. C. (1983). The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *The Journal of Neuroscience*, *3*, 2563–2586.

Maunsell, J. H., & van Essen, D. C. (1987). Topographic organization of the middle temporal visual area in the macaque monkey: representational biases and the relationship to callosal connections and myeloarchitectonic boundaries. *Journal of Comparative Neurology*, *266*(4), 535–555.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*, 746–748.

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*, 640–662.

Metea, M. R., & Newman, E. A. (2006). Glial cells dilate and constrict blood vessels: a mechanism of neurovascular coupling. *The Journal of neuroscience*, *26*(11), 2862–2870.

Meyer, G. F., & Wuerger, S. M. (2001). Cross-modal integration of auditory and visual motion signals. *NeuroReport*, *12*(11), 2557–2560.

Middlebrooks, J. C., & Pettigrew, J. D. (1981). Functional classes of neurons in primary auditory cortex of the cat distinguished by sensitivity to sound location. *J Neurosci*, *1*(1), 107–120.

Miller, M. W., & Vogt, B. A. (1984). Direct connections of rat visual cortex with sensory, motor, and association cortices. *The Journal of Comparative Neurology*, *226*, 184–202.

Morgan, L. K., MacEvoy, S. P., Aguirre, G. K., & Epstein, R. A. (2011). Distances between Real-World Locations Are Represented in the Human Hippocampus. *Journal of Neuroscience*, *31*, 1238–1245.

Morgan, M., DeAngelis, G., & Angelaki, D. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. *Neuron*, *59*(4), 662.

Morrone, M., Tosetti, M., Montanaro, D., Fiorentini, A., Cioni, G., & Burr, D. (2000). A cortical area that responds specifically to optic flow, revealed by fMRI. *Nature neuroscience*, *3*(12), 1322–1328.

Mountcastle, V., et al. (1957). Modality and topographic properties of single neurons of cats somatic sensory cortex. *J. Neurophysiol*, *20*(4), 408–434.

Mumford, J., Turner, B., Ashby, F., & Poldrack, R. (2012). Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. *NeuroImage*, *59*(3), 2636–2643.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of Cue Integration in Human Navigation. *Current Biology*, *18*(9), 689–693.

Naselaris, T., Kay, K., Nishimoto, S., & Gallant, J. (2011). Encoding and decoding in fMRI. *Neuroimage*, *56*(2), 400–410.

Nath, A., & Beauchamp, M. (2011). Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech. *The Journal of Neuroscience*, *31*(5), 1704–1714.

Nauhaus, I., Busse, L., Carandini, M., & Ringach, D. (2008). Stimulus contrast modulates functional connectivity in visual cortex. *Nature neuroscience*, *12*(1), 70–76.

Newsome, W. T., Wurtz, R. H., & Komatsu, H. (1988). Relation of cortical areas MT and MST to pursuit eye movements. II. Differentiation of retinal from extraretinal inputs. *Journal of Neuroscience*, *60*, 604–620.

Ohshiro, T., Angelaki, D., & DeAngelis, G. (2011). A normalization model of multisensory integration. *Nature neuroscience*, *14*(6), 775–782.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*, 171–175.

Orban, G. A. (2008). Higher Order Visual Processing in Macaque Extrastriate Cortex. *Physiological Review*, *88*, 59–89.

Page, W. K., & Duffy, C. J. (2003). Heading representation in MST: sensory interactions and population encoding. *Journal of Neurophysiology*, *89*(4), 1994–2013.

Parkkonen, L., Andersson, J., Hämäläinen, M., & Hari, R. (2008). Early visual brain areas reflect the percept of an ambiguous scene. *Proceedings of the National Academy of Sciences*, *105*(51), 20500–20504.

Parks, T. E. (1965). Post-retinal visual storage. *The American journal of psychology*, *78*(1), 145–147.

Penhune, V. B., Zatorre, R. J., MacDonald, J., & Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cerebral Cortex*, *6*, 661–672.

Perrott, D., & Musicant, A. (1977). Rotating tones and binaural beats. *The Journal of the Acoustical Society of America*, *61*, 1288.

Pitzalis, S., Sereno, M. I., Committeri, G., Fattori, P., Galati, G., Patria, F., & Galletti, C. (2010). Human v6: the medial motion area. *Cerebral Cortex*, *20*(2), 411–424.

Poirier, C. (2005). specific activation of the V5 brain area by auditory motion processing: An fMRI study. *Cognitive Brain Research*, *25*, 650–658.

Pouget, A., Deneve, S., Ducom, J.-C., & Latham, P. E. (1999). Narrow versus wide tuning curves: What's best for a population code? *Neural Computation*, *11*(1), 85–90.

Pouget, A., Deneve, S., Duhamel, J., et al. (2002). A computational perspective on the neural basis of multisensory spatial representations. *Nature Reviews Neuroscience*, *3*(9), 741–747.

Prokop, T., Schubert, M., & Berger, W. (1997). Visual influence on human locomotion modulation to changes in optic flow. *Experimental Brain Research*, *114*(1), 63–70.

Putzar, L., Goerendt, I., Lange, K., Rösler, F., & Röder, B. (2007). Early visual deprivation impairs multisensory interactions in humans. *Nature neuroscience*, *10*, 1243–1245.

Rademacher, J., Morosan, P., Schleicher, A., Freund, H. J., & Zilles, K. (2001). Human primary auditory cortex in women and men. *NeuroReport*, *12*, 1561–1565.

Raemaekers, M., Lankheet, M., Moorman, S., Kourtzi, Z., & van Wezel, R. (2009). Directional anisotropy of motion responses in retinotopic cortex. *Human brain mapping*, *30*(12), 3970–3980.

Rauschecker, J. P., & Tian, B. (2000). Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proceedings of the National Academy of Sciences, 97*, 11800–11806.

Riecke, B. E., Feuerissen, D., & Reiser, J. J. (2009). Auditory self-motion simulation is facilitated by haptic and vibrational cues suggesting the possibility of actual motion. *ACM Transactions on Applied Perception, 6*(3), 20:1–20:22.

Roach, N., Heron, J., & McGraw, P. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings of the Royal Society B: biological sciences, 273*(1598), 2159–2168.

Rodríguez, G. (2012). Lecture notes on generalized linear models. Chapter 5: Log-Linear Models for Contingency Tables.

Rodriguez, P. (2010). Neural decoding of goal locations in spatial navigation in humans with fmri. *Human brain mapping, 31*(3), 391–397.

Roe, A. W., Fritsches, K., & Pettigrew, J. D. (2005). Optical imaging of functional organization of v1 and v2 in marmoset visual cortex. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology*, *287*(2), 1213–1225.

Romanski, L. M. (2007). Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cerebral Cortex*, *17*, i61–i69.

Rosa, M. G., Soares, J. G., Fiorani, M., & Gattass, R. (1993). Cortical afferents of visual area mt in the cebus monkey: possible homologies between new and old world monkeys. *Visual neuroscience*, *10*(5), 827–855.

Sakamoto, S., Osada, Y., Suzuki, Y., & Gyoba, J. (2004). The effects of linearly moving sound images on self-motion perception. *Acoustical Science and Technology*, *25*(1), 100–102.

Schaafsma, S., Duysens, J., Gielen, C., et al. (1997). Responses in ventral intraparietal area of awake macaque monkey to optic flow patterns corresponding to rotation of planes in depth can be explained by translation and expansion effects. *Visual neuroscience*, *14*(4), 633–646.

Scheef, L., Boecker, H., Daamen, M., Fehse, U., Landsberg, M. W., Granath, D., Mechling, H., & Effenberg, A. O. (2009). Multimodal motion processing in area V5/MT: Evidence from an artificial class of audio-visual events. *Brain Research*, *1252*, 94–104.

Schipke, C. G., & Kettenmann, H. (2004). Astrocyte responses to neuronal activity. *Glia*, *47*(3), 226–232.

Schlack, A., Hoffmann, K. P., & Bremmer, F. (2003). Selectivity of macaque ventral intraparietal area (area VIP) for smooth pursuit eye movements. *Journal of Physiology*, *551*(2), 551–561.

Schlack, A., Sterbing-D'Angelo, S., Hartung, K., Hoffmann, K., & Bremmer, F. (2005). Multisensory space representations in the macaque ventral intraparietal area. *The Journal of neuroscience*, *25*(18), 4616–4625.

Schluppeck, D., Glimcher, P., & Heeger, D. J. (2005). Topographic organization for delayed saccades in human posterior parietal cortex. *Journal of Neuroscience*, *94*, 1372–1384.

Schmah, T., Hinton, G., Zemel, R., Small, S., & Strother, S. (2009). Generative versus discriminative training of RBMs for classification of fMRI images. In *Proc. NIPS*, vol. 2009.

Schmiedchen, K., Freigang, C., Nitsche, I., & Rübsamen, R. (2012). Crossmodal interactions and multisensory integration in the perception of audio-visual motion - A free-field study. *Brain Research*, *1466*, 99–111.

Schönwiesner, M., Krumbholz, K., Rübsamen, R., Fink, G. R., & Cramon, Y. (2007). Hemispheric asymmetry for auditory processing in the human auditory brain Stem, thalamus, and cortex. *Cerebral Cortex*, *17*, 492–499.

Scott-Samuel, N., & Magapu, J. (2002). Vertical motion looks faster than horizontal motion. In *Perception 31*. European Conference on Visual Perception.

Scoville, W. B., & Milner, B. (2000). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neuropsychiatry Clincal Neuroscience*, *12*(1), 103–113.

Sereno, M. I., & Huang, R. S. (2006). A human parietal face area contains aligned head-centered visual and tactile maps. *Nature Neuroscience*, *9*, 1337–1343.

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

Sestieri, C., Di Matteo, R., Ferretti, A., Del Gratta, C., Caulo, M., Tartaro, A., Olivetti Belardinelli, M., & Romani, G. (2006). "What" versus "Where" in the audiovisual domain: An fMRI study. *Neuroimage*, *33*, 672–680.

Shams, L., & Beierholm, U. R. (2010). Causal inference in perception. *Trends in cognitive sciences*, *14*(9), 425–432.

Shapleske, J., Rossell, S. L., Woodruff, P. W. R., & David, A. S. (1999). The planum temporale: a systematic, quantitative review of its structural, functional and clinical significance. *Brain Research Reviews*, *29*, 26–49.

Sharp, P. E., Blair, H. T., & Cho, J. (2001). The anatomical and computational basis of the rat head-direction cell signal. *Trends in Neuroscience*, *24*(5), 289–294.

Shenoy, K., Crowell, J., & Andersen, R. (2002). Pursuit speed compensation in cortical area MSTd. *Journal of neurophysiology*, *88*(5), 2630–2647.

Shmuel, A., & Grinvald, A. (1996). Functional organization for direction of motion and its relationship to orientation maps in cat area 18. *The Journal of Neuroscience*, *16*(21), 6945–6964.

Silver, M., & Kastner, S. (2009). Topographic maps in human frontal and parietal cortex. *Trends in cognitive sciences*, *13*(11), 488–495.

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

Smiley, J., & Falchier, A. (2009). Multisensory connections of monkey auditory cerebral cortex. *Hearing research*, *258*, 37.

Smith, A., Wall, M., Williams, A., & Singh, K. (2006). Sensitivity to optic flow in human cortical areas MT and MST. *European Journal of Neuroscience*, *23*(2), 561–569.

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

Smith, K., Okada, K., Saberi, K., & Hickok, G. (2004). Human cortical auditory motion areas are not motion selective. *Neuroreport*, *15*(9), 1523–1526.

Smith, K., Saberi, K., & Hickok, G. (2007). An event-related fMRI study of auditory motion perception: No evidence for a specialized cortical system. *Brain research*, *1150*, 94–99.

Smith, M. (2009). Is the postsubiculum involved in associating landmarks with location? *Informatics MSc thesis*.

Smith, M., Wutte, M., Flanagin, V., Glasauer, S., & Wolbers, T. (2013). Correlation of brain responses to opposing directions of visual motion: A sign of columnar organisation. *in submission*.

Snowden, R., Treue, S., Erickson, R., & Andersen, R. (1991). The response of area MT and V1 neurons to transparent motion. *The Journal of Neuroscience*, *11*(9), 2768–2785.

Soames, R., & Raper, S. (1992). The influence of moving auditory fields on postural sway behaviour in man. *European journal of applied physiology and occupational physiology*, *65*(3), 241–245.

Sohlberg, M., Todis, B., Fickas, S., Hung, P., & Lemoncello, R. (2005). A profile of community navigation in adults with chronic cognitive impairments. *Brain Injury*, *19*(14), 1249–1259.

Spence, C., Senkowski, D., & Röder, B. (2009). Crossmodal Processing. *Exp Brain Res*, *198*, 107–111.

Spiers, H., & Maguire, E. (2007). A navigational guidance system in the human brain. *Hippocampus*, *17*(8), 618–626.

Stackman, R. W., & Taube, J. S. (1998). Firing properties of rat lateral mammillary single units: Head direction, head pitch, and angular head velocity. *The Journal of Neuroscience*, *18*(21), 9020–9037.

Stanford, T. R., & Stein, B. E. (2007). Superadditivity in multisensory integration: putting the computation in context. *Neuroreport*, *18*, 787–792.

Stephan, K., Penny, W., Daunizeau, J., Moran, R., & Friston, K. (2009). Bayesian model selection for group studies. *Neuroimage*, *46*(4), 1004–1017.

Stephan, K., Penny, W., Moran, R., Den Ouden, H., Daunizeau, J., & Friston, K. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage*, *49*(4), 3099–3109.

Strange, B., Fletcher, P., Henson, R., Friston, K., & Dolan, R. (1999). Segregating the functions of human hippocampus. *Proceedings of the National Academy of Sciences*, *96*(7), 4034–4039.

Stringer, S. M., Trappenberg, T. P., Rolls, E. T., & de Araujo, I. E. T. (2002). Self-organizing continuous attractor networks and path integration: one-dimensional models of head direction cells. *Network: Computation in Neural Systems*, *13*, 217–242.

Tanaka, K., & Saito, H.-A. (1989). Analysis of motion of the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey. *Journal of Neurophysiology*, *62*(3), 626–641.

Tassinari, H., Hudson, T., & Landy, M. (2006). Combining priors and noisy visual cues in a rapid pointing task. *The Journal of neuroscience*, *26*(40), 10154–10163.

Taube, J. (1995). Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. *The Journal of neuroscience*, *15*(1), 70–86.

Taube, J., Goodridge, J., Golob, E., Dudchenko, P., & Stackman, R. (1996). Processing the head direction cell signal: a review and commentary. *Brain Research Bulletin*, *40*(5), 477–484.

Taube, J. S., Muller, R. U., & Ranck, J. J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *The Journal of Neuroscience*, *10*(2), 420–435.

Tetewsky, S., & Duffy, C. (1999). Visual loss and getting lost in alzheimers disease. *Neurology*, *52*(5), 958–958.

Thompson, P. (1880). Optical illusions of motion. *Brain*, *3*(3), 289–298.

Tootell, R. B., et al. (1998). Functional analysis of primary visual cortex (V1) in humans. *Proceedings of the National Academy of Sciences*, *95*, 811–817.

Urbantschitsch, V. (1897). Ueber Störungen des Gleichgewichtes und Scheinbewegungen. *Zeitschrift für Ohrenheilkunde, Band XXXI*, 234–294.

Väljamäe, A., Larsson, P., Västfjäll, D., & Kleiner, M. (2008). Sound representing self-motion in virtual environments enhances linear vection. *Presence: Teleoperators and Virtual Environments*, *17*(1), 43–56.

Väljamäe, A., et al. (Eds.) (2004). *Auditory Presence, Individualized Head-Related Transfer Functions, and Illusory Ego-Motion in Virtual Environments*, 18536. International Society for Presence Research.

Van den Berg, A. (1992). Robustness of perception of heading from optic flow. *Vision research*, *32*(7), 1285–1296.

Vann, S., Aggleton, J., & Maguire, E. (2009). What does the retrosplenial cortex do? *Nature Reviews Neuroscience*, *10*(11), 792–802.

Vouloumanos, A., Kiehl, K. A., Werker, J. F., & Liddle, P. F. (2001). Detection of sounds in the auditory stream: event-related fMRI evidence for differential activation to speech and nonspeech. *Journal of Cognitive Neuroscience*, *13*, 994–1005.

Wall, M., & Smith, A. T. (2008). The representation of egomotion in the human brain. *Current Biology*, *18*, 191–194.

Wandell, B., Brewer, A., & Dougherty, R. (2005). Visual field map clusters in human cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 693–707.

Wandell, B., Dumoulin, S., & Brewer, A. (2007). Visual field maps in human cortex. *Neuron*, *56*(2), 366–383.

Wang, R. F., & Spelke, E. S. (2002). Human spatial representation: Insights from animals. *Trends in cognitive sciences*, *6*(9), 376–382.

Warren, W. H., Morris, M. W., & Kalish, M. (1988). Perception of translational heading from optical flow. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(4), 646.

Watson, A., & Pelli, D. (1983). QUEST: A bayesian adaptive psychometric method. *Attention, Perception, & Psychophysics*, *33*(2), 113–120.

Weliky, M., Bosking, W. H., & Fitzpatrick, D. (1996). A systematic map of direction preference in primary visual cortex. *Nature*, *379*, 22.

Wetzels, R., Raaijmakers, J., Jakab, E., & Wagenmakers, E. (2009). How to quantify support for and against the null hypothesis: A flexible winbugs implementation of a default bayesian t test. *Psychonomic bulletin & review*, *16*(4), 752–760.

Williams, M., Dang, S., & Kanwisher, N. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nature neuroscience*, *10*(6), 685–686.

Wilson, J. L., Jenkinson, M., & Jezzard, P. (2002). Optimization of static field homogeneity in human brain using diamagnetic passive shims. *Magnetic resonance in medicine*, *48*(5), 906–914.

Wohlgemuth, A. (1911). *On the after-effect of seen movement*. 1. Cambridge University Press.

Wolbers, T., Zahorik, P., & Giudice, N. (2011). Decoding the direction of auditory motion in blind humans. *Neuroimage*, *56*, 681–687.

Wraga, M., Shephard, J., Church, J., Inati, S., & Kosslyn, S. (2005). Imagined rotations of self versus objects: an fMRI study. *Neuropsychologia*, *43*(9), 1351–1361.

Wuerger, S., Hofbauer, M., & Meyer, G. (2003). The integration of auditory and visual motion signals at threshold. *Attention, Perception, & Psychophysics*, *65*(8), 1188–1196.

Wutte, M., Smith, M., Flanagin, V., & Wolbers, T. (2011). Physiological signal variability in hMT+ reflects performance on a direction discrimination task. *Frontiers in psychology*, *2*.

Xu, X., Collins, C., Kaskan, P., Khaytin, I., Kaas, J., & Casagrande, V. (2004). Optical imaging of visually evoked responses in prosimian primates reveals conserved features of the middle temporal visual area. *PNAS*, *101*(8), 2566–2571.

Yacoub, E., Harel, N., & Uğurbil, K. (2008). High-field fMRI unveils orientation columns in humans. *Proceedings of the National Academy of Sciences*, *105*(30), 10607–10612.

Yan, T., Jin, F., He, J., & Wu, J. (2011). Development of a wide-view visual presentation system for visual retinotopic mapping during functional MRI. *Journal of Magnetic Resonance Imaging*, *33*, 441–447.

Yan, T., & Wu, J. (Eds.) (2010). *Wide-field retinotopy distinguish subregions of the human MT+ by functional magnetic resonance imaging*.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of Cognitive Neuroscience*, *20*(1), 1–19.

Zarahn, E., Aguirre, G. K., & D'Esposito, M. (1997). Empirical analyses of BOLD fMRI statistics. *Neuroimage*, *5*, 179–197.

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

Zhang, T., & Britten, K. H. (2004). Clustering of selectivity for optic flow in the ventral intraparietal area. *NeuroReport*, *15*, 1941–1945.

Zhang, T., & Britten, K. H. (2010). The responses of VIP neurons are sufficiently sensitive to support heading judgments. *Journal of neurophysiology*, *103*(4), 1865–1873.

Zimmermann, J., Goebel, R., De Martino, F., van de Moortele, P., Feinberg, D., Adriany, G., Chaimow, D., Shmuel, A., Uğurbil, K., & Yacoub, E. (2011). Mapping the organization of axis of motion selective features in human area MT using high-field fMRI. *PloS one*, *6*(12), e28716.

Zvyagintsev, M., Nikolaev, A., Thönnessen, H., Sachs, O., Dammers, J., & Mathiak, K. (2009). Spatially congruent visual motion modulates activity of the primary auditory cortex. *Experimental brain research*, *198*, 391–402.