# THE NEURAL BASIS OF ATTENTION AND PERCEPTION IN THE HUMAN BRAIN

**Susanne Watkins**

Institute of Cognitive Neuroscience
Wellcome Trust Centre for Neuroimging
Institute of Neurology
University College London

Prepared under the supervision of:
Professor Geraint Rees
Professor Nilli Lavie

Submitted to UCL for the Degree of PhD

**DECLARATION**

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

The work presented in chapters 3, 5 & 6 has been published in the following papers.

Watkins,S., Dalton,P., Lavie,N., and Rees,G. (2007). Brain mechanisms mediating auditory attentional capture in humans. Cereb.Cortex *17*, 1694-1700.

Watkins,S., Shams,L., Tanaka,S., Haynes,J.D., and Rees,G. (2006). Sound alters activity in human V1 in association with illusory visual perception. Neuroimage. *31*, 1247-1256.

Watkins S, Shams,L., Josephs O, and Rees,G. (2007). Activity in human V1 follows multisensory perception. Neuroimage *17*, 572-578.

## ACKNOWLEDGEMENTS

Above all, I would like to thank my supervisors Geraint Rees and Nilli Lavie for their encouragement, guidance and advice.

I would also like to thank the members of the Rees lab – David, Richard, John, Bahador, Davina, Rimona, Elaine, Claire, Christian and Ayshe for all their help and advice over the last few years. In addition without the input of many people at the ICN including Eliot, Steffan, and Christian my time there would have been alot less productive. Thanks also to all the ICN and FIL support staff – Michelle, Karen, Dominic, Marcia, Amanda, Jan, David, Ric, Rachel, Chris, Lambert and Martin for all their help. Special thanks to Eric and Oliver, for providing much needed practical support in overcoming an array of technical issues.

I would also like to thank my family and friends for their moral support and willingness to lie in the scanner for hours of testing.

Thanks also to the Wellcome Trust for funding me.

## ABSTRACT

Being able to focus on the task at hand while retaining the ability to respond to salient task-irrelevant stimuli is critical to successful human behaviour. It is vital that animals and people can quickly redirect their attention when faced with novel or potentially threatening stimuli. In this thesis I use a range of fMRI techniques including retinotopic mapping and multivariate analysis to investigate the behavioural and perceptual consequences of task irrelevant stimuli in audition and vision. Initially I describe two fMRI experiments investigating the cortical areas mediating behaviourally defined attentional capture by a task-irrelevant auditory and visual stimulus. I then go on to demonstrate that task irrelevant auditory stimuli can have a profound effect on both visual perception and processing in early visual cortical areas. In particular I demonstrate for the first time that an auditory induced change in visual perception can influence processing in the primary visual cortex. Further more, I demonstrate that auditory timing can alter the perceived direction of visual apparent motion and that such behavioural changes can be decoded from V3 and MT+. Finally I demonstrated that in the situation where an auditory stimulus has no behaviour or perceptual relevance to visual processing early visual areas do not encode information about the auditory stimulus. Taken together these findings indicate that task irrelevant distractors can have a significant effect on behaviour and perception.

**CONTENTS**

## LIST OF FIGURES

## LIST OF TABLES

# CHAPTER 1: GENERAL INTRODUCTION

## 1.1 Introduction

A typical scene contains far too much information for the human brain to process simultaneously. In order to use information encoded in sensory signals to effectively guide behaviour, some kind of selective mechanism is required. Attention is the name given to the process of selecting the 'important' aspects of a scene for further processing, while relegating the rest to limited analysis. Attention plays an important role in our perception of the world. It is remarkable how little we perceive when we do not pay attention.

## 1.2 Attention

### *1.2.1 Spatial attention*

Humans can choose to actively attend to a particular location in space. This act of voluntary spatial attention enhances the processing of stimuli at that location. Studies of spatial attention typically require subjects to focus attention on a small part of the scene and report information at the focus of attention. For example, in a visual cueing paradigm, subjects are required to respond as quickly as possible to the onset of a light or other simple visual stimulus. This target stimulus is preceded by a "cue" whose function is to draw attention to the occurrence of a target in space. Cues come in various forms, e.g. a symbol, like an arrow, indicating where attention should be deployed (Jonides and Irwin, 1981) (Figure 1.1 a). In this case, spatial attention is deployed voluntarily to the cued location and this facilitates detection of and response to stimuli presented at the cued location (Cheal and Gregory, 1997;Luck et al., 1996). However, cueing can also be involuntary and driven by 'bottom up' factors such as the brightening of the location where the cued object will subsequently appear (Posner and Cohen, 1984) (Figure 1.1 b). There is evidence that the spatial attention system is supramodal in the sense that crossmodal spatial (tactile or vision) cues will enhance the processing for auditory objects at that location (Hotting et al., 2003). A common

analogy is to describe attention as a spotlight that enhances the efficiency of the detection of events within its beam.

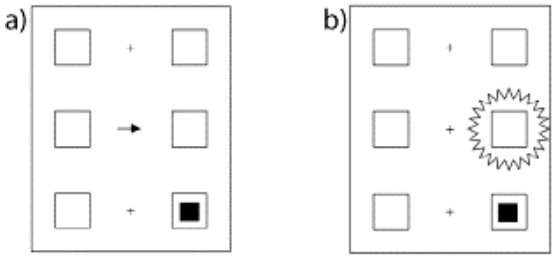


**Figure 1.1 Spatial cueing paradigm**

In Posner's early experiments the attentional cue could be one of two types; either a central endogenous cue such as an arrow pointing towards a possible target (a) or a peripheral exogenous cue such as a flash of light in the position of a target location (b).

### *1.2.2 Non-spatial attention*

Humans can also attend to individual target objects and ignore distractor objects even when they are overlapping in space (O'Craven et al., 1999) (Figure 1.2). For example, Neisser and Becklen (Neisser and Becklen, 1975) presented two different movie sequences that entirely overlapped with each other in space. Subjects were asked to attend to one of the two overlapping movies. Throughout viewing, subjects were able to follow actions in the attended movie. Odd events in the unattended movie were rarely noticed. Because both scenes overlapped each other, this demonstrates that selective attention cannot be purely space-based. Rather attentional selection was based on objects and events. Attention can also be directed to an individual feature of the visual scene such as colour, shape or direction of motion. In the auditory domain previous research has established that auditory attention, like visual attention, can focus on stimuli containing a particular (auditory) feature. For example, early studies using the dichotic listening technique found that participants could selectively attend to a channel defined by a certain auditory feature (e.g., words spoken by a female voice) while apparently ignoring the channel that did not share that feature (e.g., words spoken by a male voice; Cherry, 1953; Moray, 1959). Feature based attention may well be more important than spatial attention in the auditory domain. Previous

research suggests that the auditory system, unlike the visual system, processes spatial location with a lower priority than other stimulus features (Kubovy, 1981).

The extent to which spatial, object-based and feature-based attention share common neural mechanisms remains an open question.

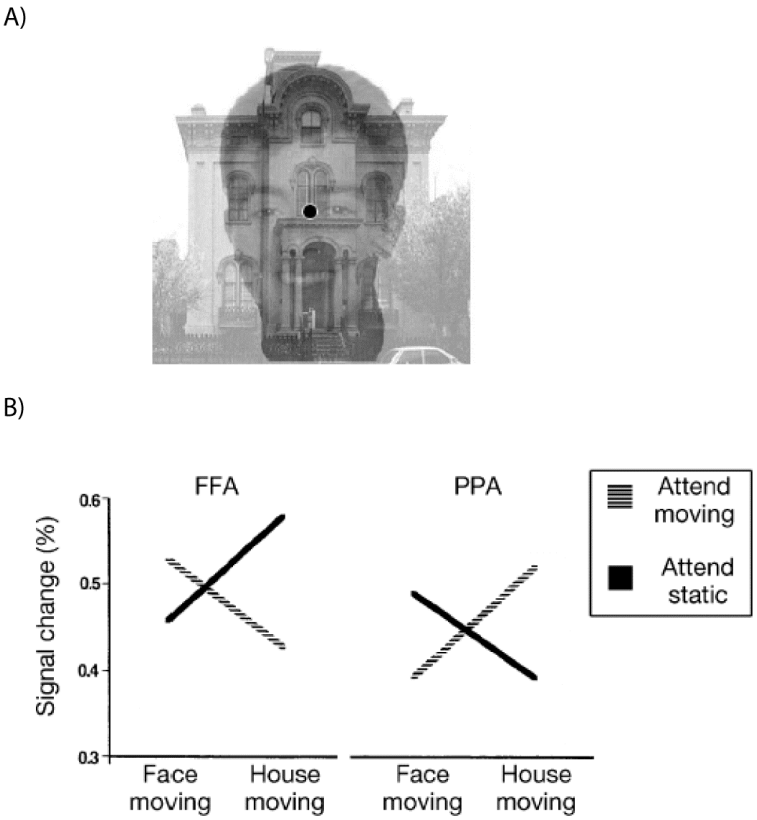


**Figure 1.2. Object-based attention.**

Subjects fixate on the central circle. A) Subjects are asked to attend to either the face or the house. B) Cortical activity measured using fMRI in the fusiform face area (FFA) and the parahippocampal place area (PPA). The FFA and PPA are functionally defined cortical areas in the ventral visual stream that respond selectively to faces rather than objects (FFA), or objects rather than faces (PPA). When subjects attend to a face, activity is higher in the FFA compared to when they attend to a house and when subjects attend to a house activity is higher in PPA than when they attend to a face. Subjects are able to differentially attend to faces or houses even when they overlap in space (adapted from O'Craven et al 1999)

In addition to varying across space, stimuli can also change rapidly with time. People need to be able to extract behaviourally relevant information from this rapid flux at

particular times. Attention can be deployed at different moments in time to facilitate information processing. For example, if a stream of visual items is rapidly presented at central fixation (rapid serial visual presentation, RSVP), subjects are generally very good at temporal selection and able to accurately detect a single target even when visual items are presented at up to 25 items per second (Sperling et al., 1971). Temporal attention is likely to be very important in the auditory domain since many auditory patterns unfold in time. Previous research has demonstrated that auditory attention can be temporally directed to focus at a particular point in time (Coull et al., 2000;Lange and Roder, 2006). In addition subjects performance improved when subjects were temporally cued to expect an auditory target (Best et al., 2007).

### *1.2.3 Potential neural mechanisms underlying attention*

Directing attention to a spatial location has many behavioural advantages: improving the accuracy and speed of responses to target stimuli at that location (Posner, 1980); increasing perceptual sensitivity; reducing interference from distractors (Cheal and Gregory, 1997;Luck and Hillyard, 1994) and improving visual acuity (Carrasco et al., 2006). The neural basis of these behavioural effects is still an issue of active investigation. Attention is thought to affect neural processing in several ways: amplification of neural responses to an attended stimulus (Lu and Dosher, 1998;Treue and Maunsell, 1996); filtering of unwanted information by suppressing nearby distractors (Beck and Kastner, 2005;Luck and Hillyard, 1994); increasing baseline activity of an attended location in the absence of stimulation; and increasing the stimulus salience by enhancing the neuron's sensitivity to stimulus contrast (Lu and Dosher, 1998). It is clear from the above that attention is not one unified process. Functional MRI of human subjects shows that selective visual attention can affect cortical responses to a visual stimulus, not only as early as the lateral geniculate nucleus (O'Connor et al., 2002) or superior colliculus (Schneider and Kastner, 2009), but also further down the processing stream in striate or extrastriate cortex (Martinez et al., 1999) (Figure 1.3). Selective attention therefore operates at multiple levels in visual processing.

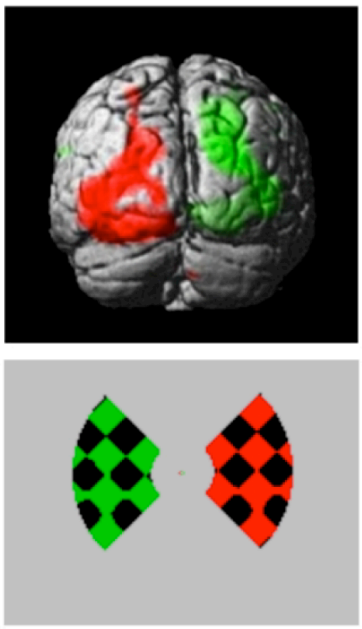

**Figure 1.3. The effect of attention in visual cortex**

fMRI data showing the effect of visual attention on striate and extrastriate cortex. A) Subjects are asked to fixate on a central cross and covertly attend to either the green or red checkerboard. B) Cortical activation is shown projected onto a 3D representation of the posterior cortical surface, showing the occipital cortex. The pattern of activation is shown in green for attending to the green checkerboard and red for the red checkerboard.

## 1.3 The control of attention

The control of visual attention reflects both cognitive (’top down’) factors such as knowledge and current goals and stimulus-driven (’bottom up’) factors that reflect the salience of sensory information. Typically, these two factors dynamically interact to control where and to what we pay attention (Corbetta & Shulman, 2002) (Figure 1.4). There is increasing evidence that two partially segregated neural networks underpin top-down control on the one hand, and bottom-up salience-driven selection on the other.

Humans are better at detecting an object in a visual scene if they have prior knowledge about its features, such as colour, motion or the time at which it will

appear (Dosher and Lu, 2000;Eriksen and Hoffman, 1973;Posner, 1980). This advance information about the object or about the location it will appear can be used to bias neurons analysing the incoming visual information in order to facilitate detecting the appropriate object in the visual scene.

Bottom-up attentional mechanisms operate on raw sensory signals, rapidly and involuntarily shifting attention to salient visual features. The attention grabbing effect of a salient stimulus can be demonstrated by flashing a light at a specific location in the visual field and measuring how long it takes for a subject to respond to a subsequent target in that location compared to another location in the visual field (Figure 1.1b). Even when this cue provides no information about the location of the forthcoming target, the cue facilitates detection and discrimination at the cued location. The facilitation produced by these 'bottom up' sensory cues is more rapid than that produced by top down cognitive cues. In addition sensory cues cause a prolonged inhibition of processing at the cued location after the early facilitation (known as 'inhibition of return'). These differences in the effects of cognitive and sensory cues have led to the idea of a functional distinction between top down and bottom up orientating systems. However, in the real world, the salience of an object is often highly dependent on its behavioural relevance. For example, if we search a crowd of people for a friend wearing a green coat then we are more likely to notice other people with green clothing. The bottom up (sensory) salience of green objects depends on the ongoing cognitive task of finding a green object.

There is some evidence that top down control of both spatial and non spatial shifts of auditory attention activate similar regions of the dorsal frontoparietal network to visual attention suggesting a crossmodal attentional network (Shomstein and Yantis, 2006).

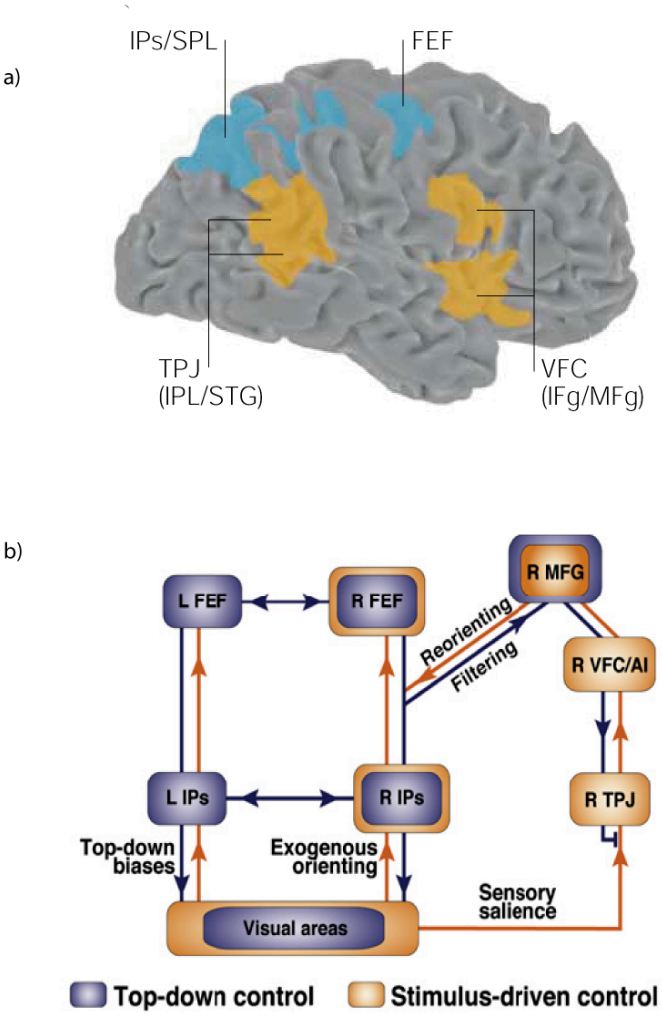


**Figure 1.4 Neuroanatomical model of attentional control.**

a) Regions in blue are consistently activated by central cues, indicating where an object will appear. Regions in orange are consistently activated when attention is reoriented to an unexpected but *behaviourally relevant* object.

b) Model for the interaction of dorsal and ventral networks during behaviourally relevant stimulus-driven reorientating. Dorsal network regions are thought to send top-down signals to visual areas, and via MFG to the ventral network (filtering signal), restricting ventral activation to behavioural relevant stimuli (adapted from Corbetta and Shulman, 2008).

## 1.4 Attentional capture.

The basic premise for attentional capture was suggested by Theeuwes in the early nineties (Theeuwes, 1992). When attention is divided across the visual field early processing is postulated to be exclusively driven by the bottom up properties of the

stimulus field. In other words attention will shift in an automatic, exogenous fashion to the location having the highest local feature contrast or salience. The salient feature is said to have captured attention.

For example, when a target in a search display contains an item that is unique on some feature (e.g. a green diamond among green circles), this salient feature appears to 'pop out' of the display, making search efficient. If however a non-target stimulus has a unique singleton feature, it will typically disrupt search performance. Such interruption of goal-driven attention can be found even when the object is a singleton on a dimension that is never relevant to the task, suggesting that attention was captured by the singleton (Figure 1.5).

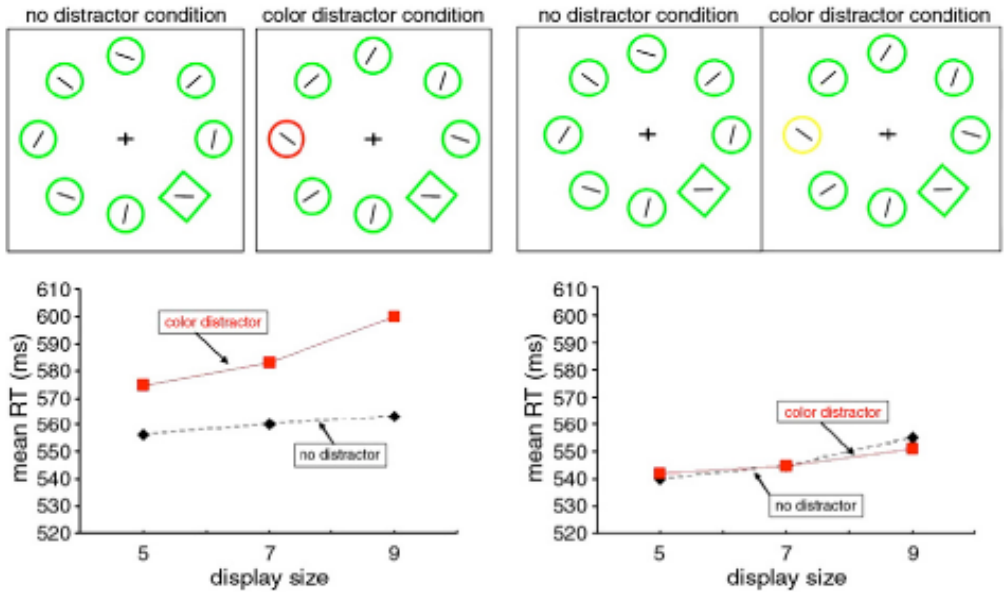


**Figure 1.5 Visual attentional capture.**
Observers search for a shape singleton, a green diamond among a variable number of green circles. Observers respond to the orientation (horizontal or vertical) of the line segment presented within the target diamond shape. When a salient colour distractor is presented (left side) it disrupts search performance and is said to have captured attention. However when the colour distractor is less salient than the target singleton, search is not effected (right side). These results are interpreted to indicate that even though observers always search for a green diamond, this top down set cannot prevent the capture of attention by a salient colour singleton. (adapted from Theeuwes (1992)).

There is on-going debate in the literature about whether feature singletons capture attention in a *purely* stimulus driven fashion. Theeuwes has presented a body of research that suggests that the first sweep of information through the brain is completely stimulus driven and based on the salience of the objects in the visual field (Theeuwes, 2010). An alternative view point proposed initially by Folk suggests that feature singletons do not automatically capture attention unless they are related to the target (i.e. the target is also a singleton) (Folk et al., 1992). Whatever the case, the body of behavioural data indicates that attention may be captured under appropriate conditions, even when the stimulus is irrelevant to the task at hand.

Recently Dalton and Lavie have established the phenomenon of auditory attentional capture (Dalton and Lavie, 2004). They designed an auditory search task in which participants were asked to search for an auditory feature target (e.g., defined by frequency) among irrelevant non-targets (with a different frequency) and to indicate whether the target was present or absent or discriminate its feature value (e.g., high frequency or low frequency). In a similar manner to the visual attentional capture paradigms, one of the non-targets could also be presented with an irrelevant singleton feature (e.g., higher intensity). Irrelevant variation in the frequency or intensity of one of the non-target tones ("distractor singleton") increased reaction times and error rates suggesting that the irrelevant feature singleton captured attention.

### *1.4.1 Brain mechanisms mediating attentional capture in humans*

Logically the ventral attentional network would be considered the likely candidate for mediating attentional capture. However this hypothesis has now been tested and rejected. Several, recent visual studies have shown that activation of the right lateralized ventral attentional network critically depends on the behavioural relevance of the particular object that captures attention (specifically, sharing a feature with the target of the search; (Kincade et al., 2005;Serences et al., 2005). By contrast, in attentional capture the distractor singleton is never behaviourally relevant to the search task. Therefore activity of the ventral network would not be anticipated.

#### *1.4.1.1 Auditory attentional capture and the mismatch negativity*

In the auditory domain many previous studies concerning the neural response to deviant auditory stimuli have concentrated on the pre-attentive process of detection of such auditory deviants. Specifically, generation of the electrophysiological potential known as the mismatch negativity (MMN) is associated with the pre-attentive detection of deviant auditory stimuli (Jaaskelainen et al., 2004;Liebenthal et al., 2003;Opitz et al., 1999b;Opitz et al., 1999a;Schroger, 1994;Wolff and Schroger, 2001). Most previous studies of the MMN required subjects to passively listen to a stream of auditory stimuli with no measure of the behavioural response to the presence of a deviant auditory stimulus. Without a concurrent behavioural measure of any distraction, such studies do not distinguish neural responses associated with acoustic variability per se from those specific to attentional capture. Recent studies have measured the electrophysiological potentials associated with a deviant auditory stimulus and related them to a reaction time measure of behavioural distraction (Berti et al., 2004;Escera et al., 1998;Escera et al., 2001;Rinne et al., 2006;Roeber et al., 2003;Schroger et al., 2000;Schroger and Wolff, 1998). These revealed that the presence of a rare deviant auditory stimulus elicited MMN, N1 and P3a ERP components and slowed reaction times in a subsequent auditory (Alho et al., 1997;Berti and Schroger, 2004;Rinne et al., 2006;Roeber et al., 2003;Schroger et al., 2000;Schroger and Wolff, 1998) or visual task (Escera et al., 1998;Escera et al., 2001). However, the presence of behavioural interference was always associated with increased auditory variability in the stimulus. Thus, such studies cannot distinguish neural responses associated with acoustic variability in the stimulus from those specific to attentional capture.

As I have discussed previously there is evidence that top down control of both spatial and non spatial shifts of auditory attention activate similar regions of the dorsal frontoparietal network to visual attention (Shomstein and Yantis, 2006). This suggests that the same control network is shared between both visual and auditory attention. It is an unresolved question as to whether auditory attentional capture activates the same cortical network as visual attentional capture.

In chapter 3 I designed an experiment to establish for the first time the cortical areas associated with behaviourally defined attentional capture in audition. Importantly, the design allowed me to distinguish cortical areas responsible for auditory change detection from those responsible for auditory attentional capture.

#### *1.4.1.2 Visual attentional capture*

De Fockert and colleagues have previously investigated the neural substrates of singleton capture in visual search (de Fockert et al., 2004). Neural activity was measured via functional magnetic resonance imaging (fMRI) as subjects viewed similar search displays to Theeuwes (1991). Subjects searched for a circle among diamonds, and on 25% of the trials the target was a colour singleton, whereas on another 25% of the trials one of the distractors was a colour singleton. Although there was no measured neural activity specifically related to the colour singleton target, the presence of a colour singleton distractor led to bilateral activation within the dorsal attentional network (bilateral parietal cortex and left frontal cortex) relative to when no colour singleton was present. This is compatible with the hypothesis that uninformative but salient distractors are associated with activation of the dorsal rather than ventral attentional network (Corbetta et al., 2008).

A recent TMS study has questioned the conclusion that bilateral PPC is involved in attentional capture. Hodsoll and colleagues performed rTMS over the left and right parietal cortex during an attentional capture task. They demonstrated that rTMS over the right parietal cortex abolished the behavioural interference effect while rTMS over the left parietal cortex had no effect. They concluded that the right PPC had a critical role in visual attentional capture.

#### *1.4.1.3 Controlling distraction from irrelevant singleton distractors*

The ability to ignore or at least suppress misleading information is vital to the successful deployment of attention. On a first glance at a scene with a salient distractor, bottom up stimulus driven mechanisms guide attention to the most salient item in the visual field. However if bottom up salience were the end of the story, one

would become the slave of stimulus and never be able to locate the target of the search. Instead, top down attentional control mechanisms can be employed to return attention to the appropriate target.
De Fockert and colleagues found a negative correlation between the magnitude of the activity in the left frontal cortex and the level of interference produced by the distractor singleton, suggesting a role for this area in the control of interference from irrelevant distractors (de Fockert et al., 2004). An additional study demonstrated that the degree of attentional capture is influenced by cognitive load, suggesting that attentional capture is subject to top down control (Lavie and de Fockert, 2005).

There has been a considerable amount of research concerning the type of stimulus that will capture attention and whether such attentional capture is mandatory. In the case of visual attentional capture there is evidence that a distractor singleton needs to be more salient than the target to capture attention (Theeuwes, 1992). However, no one has examined the effect of systematically varying the distractor salience on the behavioural interference.

In chapter 4 I investigated the neural correlates of visual attentional capture. In addition I examined the behavioural and cortical effects of systematically varying the level of salience of the distractor singleton.

Thus far, I have concentrated on the attentional effects of task irrelevant singleton distractors. In the next sequence of studies (chapters 5 - 8) I turn my attention to the effects of task irrelevant auditory stimuli on visual perception and cortical processing. I present a series of experiments where subjects were asked to perform a visual task in the presence of accompanying auditory stimuli. The auditory stimuli were never explicitly task relevant and the subjects were always asked to ignore them and only respond to the visual events. Despite these instructions the auditory stimuli caused dramatic changes to visual perception and cortical visual processing.

## 1.5 Crossmodal perception

Our perception of the world clearly benefits from the information delivered via different sensory modalities. Historically, research concerning sensory processing and perception has concentrated on one modality at a time. However in real-world situations, incoming stimuli across different modalities often arise from the same external object. Multisensory integration has often been described as occurring automatically. Early studies investigating the response properties of single neurons in anaesthestised animals demonstrated multisensory integration provided there was spatial and temporal concordance between the stimuli (Stein et al., 2004;Stein and Arigbede, 1972;Wallace et al., 1998) (see figure 1.6). Behavioural work in humans has demonstrated that crossmodal model integration can occur preattentively (Driver, 1996;Van der Burg et al., 2008). Van der Burg and colleagues investigated the influence of auditory stimuli on visual attentional capture (see section 1.4). When an auditory stimulus is presented at the same time as the visual target the search times to identify the target significantly decreased (Van der Burg et al., 2008). They proposed that the temporal information of the auditory signal is integrated with the visual signal to create a salient feature that results in attentional capture. Several more recent studies have suggested that multisensory integration may occur across various stages of stimulus processing and be subject to attentional modulation. See (Talsma et al., 2010) for a review.

Vision has previously been suggested as the dominant sensory modality and conflicting information from competing modalities was often thought to be ignored. Examples of this include the ventriloquism effect (Howard and Templeton, 1966) and visual capture (Hay et al, 1965). However, while the best know examples of crossmodal integration involve the modification of other sensory modalities by vision, there is increasing evidence that auditory stimuli can modify visual perception (particularly in temporal judgments). It has long been known that the temporal properties of a visual stimulus can be affected by an accompanying auditory stimulus (misperception of the visual events as having the temporal frequency of the apparently related auditory events) (Gebhard et al., 1959;Shipley, 1964). Shams reported that a single flash can be misperceived as two flashes if paired with two bleeps (Shams et

al., 2000). Extending this work Berger and colleagues demonstrated that when multiple sounds produce the impression of more visual events than actually occurred, visual orientation discriminations can objectively improve (even through the sounds do not provide any orientation information; indeed, the subjects are told to ignore the sounds) (Berger et al., 2003). This implies that multisensory integration can affect sensory-specific judgments.

### *1.5.1 Cortical audiovisual integration*

Evidence from single-cell studies, tracer work, and recent neuroimaging indicate numerous multisensory convergence zones in the brain (Mesulam, 1998;Wallace et al., 2004). This has now been observed for numerous cortical and subcortical regions. Subcortically, layers of the superior colliculus (SC) receive inputs from somatosensory, auditory, and visual areas (e.g., (Meredith and Stein, 1983;Stein, 1978;Stein and Arigbede, 1972) (Figure 1.6).

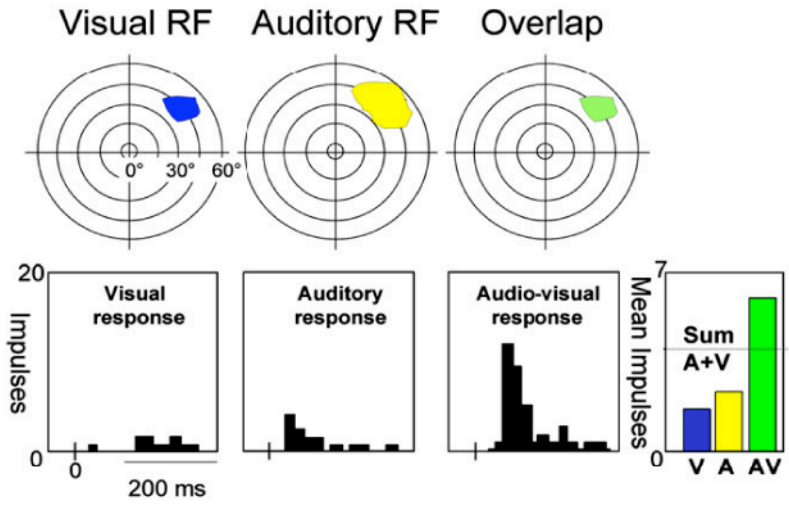


**Figure 1.6. Response properties of multisensory neurons.**

a) Response properties of a putatively illustrative multisensory neuron, in the superior colliculus, which in this case shows the nonlinearly superadditive pattern of firing. That is, the response for combined visual and auditory stimulation, with a particular spatiotemporal relation, greatly exceeds the sum of the responses to each modality alone. Adapted from (Stein et al., 2004).

In primates, the upper bank of the superior temporal sulcus is known to have bidirectional connections with unisensory auditory, visual, and somatosensory cortices (Padberg et al., 2003;Schmahmann and Pandya, 1991) and to contain multisensory neurons (Barraclough et al., 2005). Several regions within the parietal cortex are also known to receive input from sensory-specific cortices for different modalities. Finally, specific prefrontal cortical regions have also been implicated in multisensory processing (Barbas et al., 2005). Recently, some direct connections have even been reported between prefrontal cortex and primary sensory cortices (Budinger et al., 2006).

Neuroimaging studies have increasingly implicated cortical areas in multisensory integration. Shared temporal onset has been shown to activate the superior colliculus. Several regions in the superior and inferior parietal lobe appear to be involved in the integration of multisensory cues based on the shared spatial location. Finally the superior temporal sulcus (STS) has been implicated in the integration of audiovisual speech (See (Calvert and Thesen, 2004) for a review). These studies inherently assess only the more large-scale neural populations, with measures such as BOLD signal. It is therefore possible that a brain region seemingly responding to multiple modalities might comprise distinct intermixed neural populations, each responding to only one of the various senses.

### *1.5.2 Multisensory processing in unisensory areas*

Traditionally, it has been assumed that multisensory integration occurs after sensory signals have undergone extensive processing in unisensory cortical regions. However, recent studies in monkeys and humans show multisensory convergence at low-level stages of cortical sensory processing, an area previously thought to be exclusively unisensory (for a review see Foxe and Schroeder, 2005). For example, both touch and eye position can influence responses in monkey auditory association cortex and primary auditory cortex, respectively (Fu et al., 2003;Fu et al., 2004;Schroeder et al., 2001). These single unit studies are complemented by human event-related potential work demonstrating interactions between auditory and visual (Fort et al., 2002;Giard and Peronnet, 1999;Molholm et al., 2002;Molholm et al., 2004) or somatosensory

(Foxe et al., 2002;Murray et al., 2005) stimuli at very short-latency (46 -150ms). One recent fMRI study has shown evidence of audiovisual integration in BA 17 (Calvert et al., 2001) suggesting that primary visual cortex may respond to non-visual inputs. These demonstrations of early modulation of unisensory cortices by multisensory signals challenge hierarchical approaches to sensory processing (Felleman and Van Essen, 1991;Schroeder and Foxe, 2005) but the function of such multisensory convergence at the anatomically lowest stage of cortical processing remains unclear.

In chapter 5-8 of this thesis I examine the effect of task irrelevant auditory stimuli on visual perception and cortical processing. As I have discussed above there are several previous studies that have suggested that crossmodal integration occurs automatically, even when unwanted. Consistent with this view I demonstrate a dramatic change in visual perception when the visual stimulus is accompanied by an irrelevant auditory stimulus. My main interest in chapters 5-8 is to investigate the neural process underlying this perceptual change. In particular, whether this crossmodal change in visual perception is represented in primary visual cortical areas.

### 1.6 Summary of thesis

The experiments presented in this thesis all attempt to characterize the behavioural and perceptual consequences of a task-irrelevant distractor. The studies can be grouped according to whether they examine the behavioural (attentional) effect of a task irrelevant distractor singleton (chapters 3 & 4) or how the presence of a task irrelevant auditory stimulus affects visual perception and cortical visual processing (chapters 5, 6, 7 & 8).

Chapter 3 presents a study that investigates for the first time the cortical substrates of attentional capture in the auditory domain. The design of this study allowed me to differentiate between the cortical response to an auditory deviant tone (MMN) and attentional capture.

Chapter 4 presents a study that aims to determine the cortical areas underlying visual attentional capture by a task irrelevant distractor singleton. The salience of the

distractor singleton (and therefore the size of the behavioural interference effect) was varied across trials. This allowed me to investigate which cortical areas may play a role in resisting distraction by a salient singleton.

Chapter 5 presents a study examining the effects of multisensory stimulation on cortical activity in primary visual cortex. In particular I was interested in addressing the question; 'does cortical activity in early visual areas follow multisensory perception, or the physically present visual stimulus?'

Chapter 6 extends the findings of the previous chapter by demonstrating that early visual cortical activity could be modulated in either direction by an auditory stimulus, and the multisensory effects could not be explained by alerting or attention.

Chapter 7 presents a study that investigates whether activity in early visual cortex is modulated by the direction of visual and audiovisual apparent motion. No studies to date have identified the distinct neural correlates of perceiving leftwards versus rightwards directions of long-range apparent motion. In this study I use multivariate pattern classification to test whether the perceived direction of visual and audiovisual apparent motion can be reliably decoded from early visual areas.

Chapter 8 presents a study that using multivariate classification in fMRI to determine whether crossmodal signals that do not result in a change in visual perception or attention are represented in primary visual areas.

### 1.7 Conclusion

The world we live in is a complex multisensory environment. In order to complete the task at hand we must be able to resist distraction. However, it is critical that novel salient, potential life threatening stimuli are able to 'capture our attention'. The first two chapters in this thesis explore the cortical mechanisms underlying such attentional capture.

In the second half of this thesis we examine the effect of task-irrelevant auditory stimuli on visual perception and processing, particularly in early visual cortex. The integration of information from different sensory systems is a fundamental characteristic of perception and cognition – qualitatively different kinds of information from the various sense organs are put together in the brain to produce a unified, coherent representation of the outside world. Traditionally, it has been assumed that the integration of such disparate information at the cortical level was the task of specialized, higher-order association areas of the neocortex. In this thesis I investigate the contrasting assumption that much of the neocortex, even primary sensory areas, are intrinsically multisensory.

# CHAPTER 2: GENERAL METHODS

## 2.1 Introduction

This chapter covers those methods that are common to the majority of the thesis. These are functional MRI (fMRI), analyses of fMRI data using statistical parametric mapping (SPM) and retinotopic mapping. Individual chapters include detailed description of specific methods, such as long range infrared eye tracking, localisation of visual area MT and multivariate analysis that are utilized in individual studies.

## *2.2 Physics of magnetic resonance imaging*

### *2.2.1 Nuclear magnetic resonance*

Nuclear magnetic resonance was discovered in 1946 (Bloch, 1946). Nuclear magnetic resonance is essentially concerned with transitions between the energy levels of a system of nuclear spins in a magnetic field.

Consider a nucleus with spin angular momentum **I** and magnetic moment **μ**. The magnitude of the nuclear angular momentum will be given by

$$| \mathbf{I} | = \hbar[I(I+1)]^{1/2}$$

where I is the nuclear spin number and $\hbar = h/2\pi$, where h is Planck's constant. When an external magnetic field is applied with flux density $B_0$ in the z direction. The z component of angular momentum will be given by

$$I_z = m_I \hbar$$

Where $m_I$ is the magnetic quantum number taking values ± I, ±(I-1), …., giving (2l+1) possible orientations of angular momentum.

For the proton, I = ½ and $m_I$ can take values of ±½ which correspond to spin up and spin down states.

The energy of these states is the same in the absence of a magnetic field. In the presence of a magnetic field the energy difference between the two states will be given by

$$\Delta E = \gamma \hbar B_0$$

Transitions between the two states can occur under the application of a suitable radiofrequency (RF) field (Figure 2.1). i.e. one whose photons have an energy exactly equal to the energy difference between the states.

$$\hbar\omega = \gamma \hbar B_0$$

$$\omega = \gamma B_0$$

This is the Larmor equation, it is fundamental to NMR and relates the frequency of the RF signals that can be absorbed or admitted by the protons.

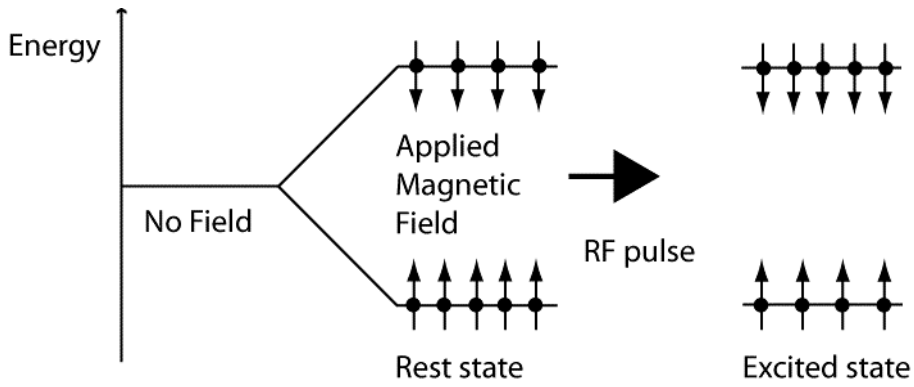


**Figure 2.1 Spin distribution**

The distribution of spin ½ nuclei (protons) between the two energy states; the rest state and the excited state (after a 180° RF pulse).

In the human brain there are many millions of protons; after all, we are about 80% water. Therefore, although an individual proton's behaviour can be predicted using quantum mechanics, the average behaviour is better explained using simple classical mechanics.

Consider a large number of N non-interacting protons. At equilibrium the distribution between the higher and lower energy states will be described by the Boltzmann distribution.

$$n_h / n_l = \exp \Delta E/KT$$

Where $n_h$ is the number of protons in the higher energy state (spin down), $n_l$ is the number of protons in the lower energy state (spin up), k represents Boltzmann's constant and T is temperature.

There is a small excess of spins in the low energy state and this small excess of spinning hydrogen atoms gives rise to a net magnetisation vector **M**.

When sufficient numbers of spins are considered it is valid to treat this net magnetisation vector classically. Consider a group of nuclear spins with a net angular momentum **A** and a net magnetisation moment **M**. **M** and **A** are related by the gyromagnetic ratio ($\mathbf{M} = \gamma\mathbf{A}$). The magnetic field $\mathbf{B_0}$ will exert a torque on the vector **M** given by

$$\mathbf{T} = \mathbf{M} \wedge \mathbf{B_0}$$

This torque will be equal to the rate of change of angular momentum of the spin system. Therefore the equation of motion is given by

$$d\mathbf{M}/dt = \gamma(\mathbf{M} \wedge \mathbf{B_0})$$

This equation describes the precession of the net magnetisation vector **M** around the magnetic field $\mathbf{B_0}$ with angular velocity $\boldsymbol{\omega} = \gamma\mathbf{B_0}$. If a magnetic field $\mathbf{B_1}$ orientated in the xy plane and rotating at the Larmor frequency is applied, **M** will experience a

torque that will rotate it into the xy plane. The angle that **M** moves through will depend on the magnitude and duration of $\mathbf{B_1}$. As soon as there is an angle between **M** and the z axis, **M** will start to precess.

Thus far we have only dealt with perturbations to the net magnetisation vector. **M** will relax back to equilibrium after the RF pulse as the energy absorbed is either readmitted or lost to the lattice.

### *2.2.2 Relaxation processes*

To understand the relaxation processes it helps to consider the longitudinal and transverse processes separately. The longitudinal process involves the return of $\mathbf{M}_z$ to its equilibrium value, the transverse relaxation describes the decay of the magnetisation in the xy plane.

#### *2.2.2.1 T1 longitudinal relaxation*

Longitudinal relaxation involves spins changing state. When a 90 degree RF pulse is applied to a system of spins, energy is absorbed, causing the number of protons in each state to become equal ($\mathbf{M}_z = 0$). The spin distribution will return to equilibrium in a time characterised by T1. T1 is affected by the composition of the environment and thus is different in different tissues, which can be used to provide contrast between tissues.

#### *2.2.2.2 T2 relaxation*

Transverse relaxation describes the loss of magnetisation in the xy plane. This process does not involve any changes to the spins population of the energy levels and is reversible. Consider a volume of spins excited by an RF pulse. Now consider this population divided up into small volumes. Due to local variations in the magnetic field caused by different chemical environments, **M** will precess at slightly different speeds. In a characteristic time T2 this will tend to lead to a loss of phase coherence and hence loss of magnetisation in the xy plane. T2 can never be longer that T1.

T2 imaging usually employs a spin echo technique, in which spins are refocused to compensate for field inhomogeneities. T2* imaging (used in fMRI) is performed without refocusing and sacrifices image fidelity in order to provide additional sensitivity for T2 relaxation processes.

### *2.2.3 MRI images*

To create an image with MRI, protons have to be distinguishable on the basis of spatial location. In an MRI scanner the signal is spatially encoded by using a combination of magnetic field gradients imposed on the static magnetic field.

Initially a magnetic field gradient is applied across the subject in the z direction. This slice-select gradient causes the proton spins at different locations to resonant at different frequencies. This means an RF pulse of a certain bandwidth will only excite spins within a certain slice, and the resulting recorded signal can only contain information from this slice. To localise the frequency in the x direction a magnetic field gradient is applied during the acquisition of the data, causing spins in a lower magnetic field to readmit their absorbed energy at a lower frequency than those in a higher field. The signal amplitude at each frequency can be extracted using a Fourier transform, with the final dimension localised using a phase encoded gradient in the y direction. Discrete increases in the phase encoding gradients divide each slice into small cubes called voxels (volume elements). The protons in a single voxel experience the same frequency and phase encoding and the signal from a voxel is the sum of the signal for all the protons in that voxel.

### *2.2.4 Echo-planar imaging*

Echo planar imaging (EPI) allows rapid acquisition of whole brain images. EPI sequences acquire data from a complete slice after a single RF pulse. This means it is possible to acquire an image of a complete slice in less than 100ms. All the functional MRI (fMRI) experiments in this thesis used EPI sequences.

### *2.2.5 BOLD signal*

fMRI measures neural activity indirectly by detecting changes in regional blood flow as indicated by blood oxygenation levels. The MRI signal can be made sensitive to the oxygenation properties of blood (so called Blood Oxygenation Level Dependent contrast), because of the paramagnetic properties of haemoglobin. When haemoglobin has no oxygen bound to it, it has a net magnetic moment. When oxygen binds to the haemoglobin, this magnetic moment disappears. Therefore the 'magnetic state' of the blood reflects its level of deoxygenation (Pauling and Coryell, 1936). It is this difference in paramagnetism that allows the oxygenated state of the blood to be detected with fMRI. The paramagnetic state of deoxyhamaglobin reduces the homogeneity of the local magnetic field and therefore reduces the T2* time constant. Thus deoxyhaemaglobin produces a smaller MRI signal than oxyhaemoglobin.

This reduction in MRI signal is what underlies the BOLD effect, as blood with more deoxyhaemoglobin will produce a reduced signal relative to highly oxygenated blood. This was first demonstrated in mice (Ogawa et al., 1990b) and subsequently demonstrated in the human visual cortex by Kwong and colleagues (Kwong et al., 1992) and Ogawa and colleagues (Ogawa et al., 1990a).

Local increases in neural activity lead to an increase in glucose metabolism in the neurons and thus an increase in oxygen consumption (Hyder et al., 1997). This causes a relative deoxygenation of the blood in the surrounding blood vessels about 100ms after the onset of neuronal activity (Vanzetta and Grinvald, 1999) followed by vasodilation and an increase in blood flow to the area 500 -1000ms after the onset of neuronal activity (Villringer and Dirnagl, 1995). This increase in blood flow swiftly reverses the deoxygenation, resulting in an overall increase in the blood oxygenation level that lasts for several seconds. This overcompensation increases the ratio of oxy to deoxy haemoglobin and therefore the increase in the BOLD signal. Thus the rise in the BOLD signal during activation indicates a decrease in the concentration of deoxyhaemoglobin in the area relative to rest. The increase in BOLD contrast is delayed with respect to the underlying neuronal activity. Typically the BOLD signal peaks 4-6 seconds after the onset of the neuronal activity. The rise and subsequent

return to baseline of the BOLD signal is known as the Haemodynamic Response function.

### *2.2.6 Neuronal basis of the BOLD signal*

The specific mechanisms underlying the BOLD signal have not yet been conclusively determined. Many researchers believe that the changes in cerebral blood flow measured by fMRI corresponds to activity in the pre-synaptic axon terminal of neurons (Jueptner and Weiller, 1995). Several studies indicate that glucose consumption by neurons mainly reflects presynaptic activity at the axon terminal.

However the relationship between glucose consumption and neural activity may not be so straightforward. There is evidence for a role of astrocytes in coupling the presynaptic activity to energy consumption via the release of glutamate from the axon terminal and its uptake by the surrounding astrocytes (Magistretti and Pellerin, 1999). This uptake of glutamate requires energy and stimulates glucose uptake by the astrocytes, resulting in glycolysis and the release of lactate. This lactate may subsequently be oxidised by the adjacent neurons to meet their energy needs. Recent research has shed doubt on the view that the BOLD signal is driven by energy use in the pre-synaptic terminals. Attwell and colleagues concluded, on the basis of the measured properties of individual ion channels , that most of the energy used during neuronal activity is expended on reversing the ion movement that generate excitatory post synaptic potentials with a smaller proportion being used to reverse the ion movements that underlie action potentials (Attwell and Laughlin, 2001).

A recent highly influential study has examined how the BOLD signal relates to multiunit activity (MUA) and local field potentials (LFPs) simultaneously recorded from electrodes in monkey primary visual cortex (Logothetis et al., 2001). MUA represents the action potentials i.e. the spiking activity of multiple neurons near (100μm) the electrode tip, while LFPs are thought to be a weightless sum of the membrane potentials of the neurons surrounding the electrode tip. Such changes in membrane potential mainly reflect synaptic activity in the dendrites and soma of neurons, so LFP is thought to represent sub threshold integrative processes.

In that study, the BOLD response was found to correlate with both MUA and the LFPs. The LFPs were a slightly better predictor of the BOLD signal and therefore the authors concluded that the BOLD signal reflects the input of and processing of a given cortical area rather than its spiking output. However this may be an overstatement because although LFPs reflect mainly membrane potentials, action potentials may also contribute.

Recent research has suggested that the increase in cerebral blood flow associated with neuronal activity may not be directly related to the energy requirements of the brain but mediated via neurotransmitters (Attwell and Iadecola, 2002). It appears that the haemodynamic response may be driven by glutamate mediated signalling, leading to an influx of $Ca^{2+}$ in postsynaptic neurons. This leads to the production of nitrous oxide, adenosine, and arachidonic acid metabolites, which in turn bring about vasodilation. According to this theory the BOLD signal represents neuronal signalling rather than energy usage.

## 2.3 fMRI analysis

All fMRI data acquired in this thesis were analysed using Statistical Parametric Mapping software, SPM2, developed at the Wellcome Department of Imaging Neuroscience (http://www.fil.ion.ucl.ac.uk/spm/). SPM is a set of MATLAB subroutines that allow the preprocessing and statistical analysis of fMRI data.

### *2.3.1 Preprocessing*

#### *2.3.1.1 Spatial realignment*

Head motion during a scan will cause changes in the fMRI signal, due to movement of the head through the magnetic field. This can be a significant confound. Realignment involves applying an affine rigid-body transformation to align each scan with a reference scan (usually the first scan or the mean of all scans) and resampling the data using spline interpolation. The six parameters of the rigid body transformation, representing adjustments to pitch, yaw, roll and X, Y, Z position are estimated interactively to minimise the sum of squares. However, even after

realignment some movement related signal will persist. This is due to the non-linear effects of movement which cannot be corrected with an affine transformation. These non-linear movement related effects can be estimated and subtracted from the original data by adding movement parameters from the realignment procedure to the design matrix as regressors of no interest during the model estimation stage of the analysis (Friston et al., 1996). Despite these measures if a subject moves their head by >5mm during a single scan the data quality is generally too low to be used.

#### *2.3.1.2 Spatial normalisation*

In order to make statistical inferences across a group of subjects, their functional data needs to be in the same space. This is achievable using spatial normalization. After realignment the mean functional image is used estimate the warping parameters that map this mean image onto a standard anatomical template image. The parameters are estimated iteratively, using a Bayesian framework, to maximise the posterior probability of the parameters being correct. The estimated warp is then applied to all images. The template used for normalization is that of the Montreal Neurological Institute. The location of the voxels is expressed using an XYZ coordinate system, where the origin is the anterior commissure. The x-axis indicates the distance left (negative) and right (positive) of the sagittal plane, the y axis indicates distance posterior and anterior to the vertical plane and the z-axis indicates the distance above and below the inter-commissural line.

#### *2.3.1.3 Coregistration to T1 structural image*

In some studies, particularly chapters 5, 6 & 7, the data was not normalised to a standard template. Instead each subject's data was realigned to that subject's structural scan.

#### *2.3.1.3 Spatial smoothing*

Normalised images are spatially smoothed with a three-dimensional isotropic Gaussian kernel of 5 – 10mm FWHM. Smoothing the data is necessary to fit the assumptions of the Gaussian random field statistical model. In addition smoothing compensates for any small variations in anatomy between subjects that still exist after normalization.

### *2.3.2 Statistical parametric mapping*

#### *2.3.2.1 Basic approach*

The approach used by SPM to make statistical inferences about fMRI data is based on the conjoint use of the General Linear Model (GLM) and Gaussian random field theory. The GLM is used to estimate parameters for the experimental variables that could explain the BOLD signal time series recorded in each voxel. The resulting statistical parameters are assembled into three dimensional images - the Statistical Parametric Maps. The voxel values of the SPM are considered to be distributed according to the probabilistic behaviour of Gaussian fields and unlikely excursions of the SPM are interpreted as regionally specific effects, caused by the experimentally manipulated variables.

#### *2.3.2.2 General linear model*

For each voxel the general linear model explains the variations in the BOLD signal time series (**Y**) in terms of a linear number of experimental variables (x) plus an error term (e). The GLM can be expressed in matrix formation as:

$$\mathbf{Y} = X\boldsymbol{\beta} + \varepsilon$$

Where **Y** is a vector of signal measurements (one per image volume) at a particular voxel and **β** is a vector of the parameters estimates. X is the design matrix containing the variables (or regressors) that are thought to explain the data. The beta parameters reflect the independent contribution of each independent explanatory variable x (also

referred to as regressors) to the value of the dependent variable. The errors are assumed to be identically and normally distributed.

The regressors, which form the columns of the design matrix X are created by placing delta functions at the time points of interest and convolving this vector with the haemodynamic response function (HRF).

The HRF is modeled on the typical BOLD response to an event. The response peaks approximately 5 seconds after stimulation and is followed by an undershoot that lasts for approximately 30 seconds.

Movement parameters, calculated during realignment, can be included in the model as regressors of no interest to account for movement artifacts which have not been corrected during realignment. Temporal confounds must also be eliminated from the data. This is achieved by applying a high pass filter to the data prior to modeling, to eliminate drifts in the magnetic field and the effects of movement. A low pass filter is also applied to attempt to eliminate the effects of respiration and heart rate.

#### *2.3.2.2 Statistics*

Inferences about the relative contribution of each explanatory variable can be made by conducting T or F tests on the parameter estimates. The null hypothesis that the parameter estimates are zero can be tested by an F statistic, resulting in an SPM(F). To compare the relative contribution of one explanatory variable to another, one can contrast or subtract the parameter estimates from one another and test whether the result is zero using a t-statistic, resulting in an SPM(T).

## 2.4 Retinotopic mapping

The response properties in early visual cortex differ significantly from those in higher visual areas. There is also wide intersubject anatomical variability of early visual areas which precludes the assignment of the borders of visual areas based on sterotactically normalized conditions (Dougherty et al., 2003) (Figure 2.2).

Fortunately early visual cortical areas are retinotopically organized. This consistent organization can be used to accurately determine the boundaries between these areas (Engel et al., 1994;Sereno et al., 1995).

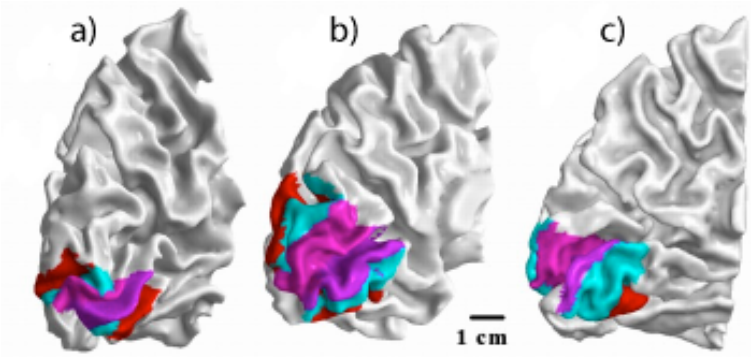


**Figure 2.2 Anatomical variability of early retinotopic visual areas across subjects.** The position and size of V1, V2 and V3 in three subjects. V1 is indicated by magenta, V2 by cyan and V3 by red. Adapted from Dougherty et al, 2003.

In order to understand the basis of retinotopic mapping, it is useful to review the anatomy of the occipital lobe. Within each hemisphere, human area V1 occupies a roughly 4cm by 8cm area located at the posterior pole of the brain in the occipital lobe. A large fraction of area V1 falls in the calcarine sulcus. From posterior to anterior cortex, the visual field representation shifts from the centre (fovea) to the periphery. The midline of V1 represents the horizontal meridian, while the boundary of V1 and V2 represents the vertical meridian. The local representation of the visual field on the cortical surface changes its orientation at the boundaries between V1 and V2 (and V2 and V3). Therefore, the spatial extent of activations elicited by visual stimuli representing the horizontal and vertical meridians can be used to functionally define these borders (Figure 2.3). This technique is called meridian mapping, and is a rapid method of retinotopic mapping. However, it provides poor information about eccentricity encoding within visual areas, and is not able to accurately define V4. To overcome these limitations usually requires the use of phase encoded retinotopic mapping methods (using a rotating wedge and expanding ring stimulus to generate a spatiotemporal pattern of stimulation of the visual field). In this thesis, meridian

mapping was used in all studies that required retinotopic mapping as the relevant experimental questions did not require accurate eccentricity information.

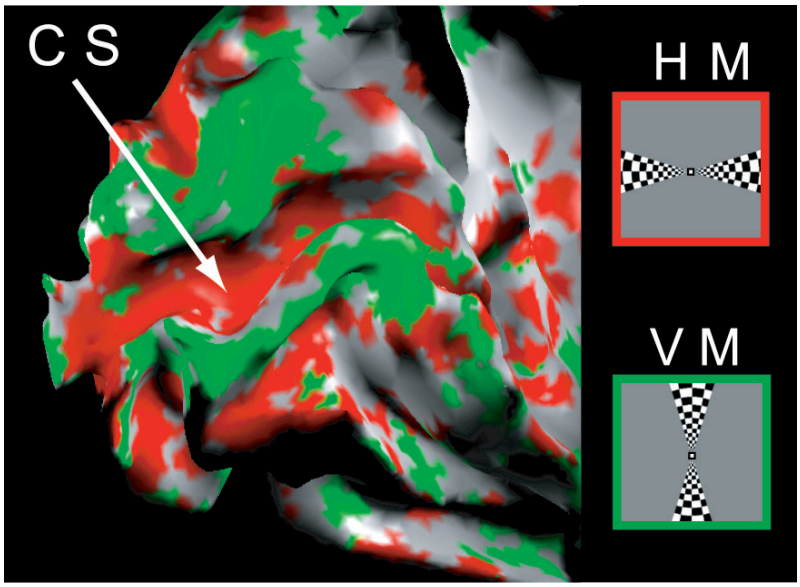


**Figure 2.3. Meridian mapping to identify cortical visual areas in human occipital cortex.** The patterns of activation elicited by horizontal (HM – shown in red) and vertical (VM – shown in green) meridian stimuli in a single subject are overlaid onto a 3D reconstruction of that subject's occipital lobe. The horizontal stimulus activates the midpoint of the calcarine sulcus (CS) and the vertical stimulus the gyri on either side of the CS. This alternating pattern of activation by horizontal and vertical meridian stimulation can be used to map the boundaries of early visual areas.

### *2.4.1 Meridan mapping*

The aim of retinotopic mapping is to accurately define the boundaries of early cortical visual areas. Initially a high resolution T1 structural scan and functional retinotopic mapping data (in response to stimuli comprising the horizontal and vertical meridians) is collected from each subject. This procedure is described in detail in the Methods section of the relevant experimental chapters of the thesis. The variation in BOLD response to these meridian stimuli is encoded in 3D Cartesian space that can be projected onto the 3D reconstruction of the anatomical image.

However, the retinotopic map is best described in terms of two-dimensional coordinates on the cortical surface, an idealized, two-dimensional representation of the cortical sheet (rather than three-dimensional Cartesian coordinates). There are two main reasons for this. First, for a given point on the cortical surface, receptive fields of neurons from different cortical layers are centered on the same point in the visual field. Second, adjacent points on the cortical surface represent adjacent points in the visual field. Therefore, retinotopic mapping first requires flattening of the cortical surface. In this thesis I used MrGray software originally developed at Stanford (Teo et al., 1997;Wandell et al., 2000) together with custom Matlab scripts developed 'in house'.

The grey and white matter in the structural scan is segmented manually (Figure 2.4). The white/grey matter that is generated during segmentation is then used to reconstruct the surface anatomy of the occipital lobe which can be represented as a mesh that nodes of grey matter can be mapped onto.

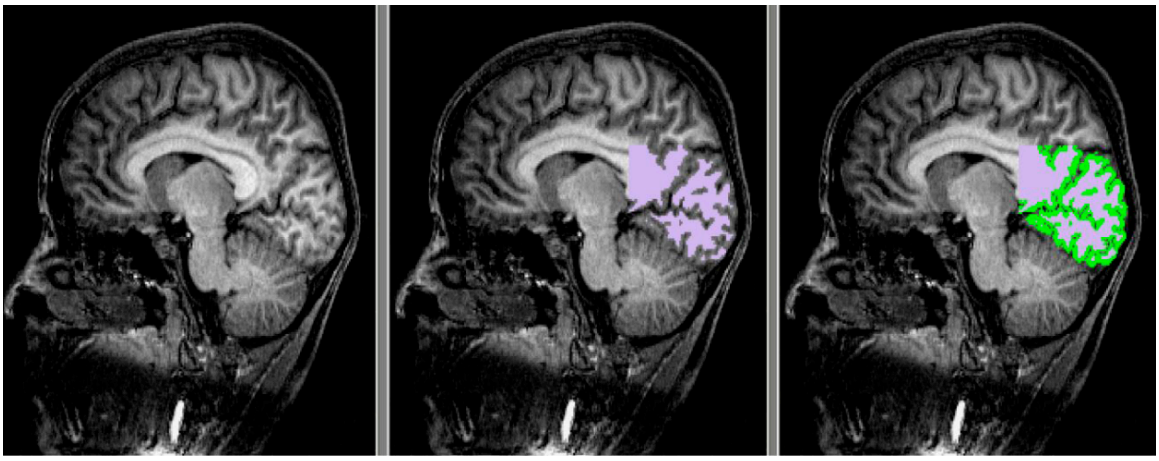

**Figure 2.4. Segmenting white and grey matter in MrGray.**
All panels show the same high resolution T1 sagittal image from a single subject. The middle panel has the white matter (shown in purple) in the occipital lobe segmented from the grey matter. The right panel has 4 layers of grey matter (shown in green) "grown" onto the white matter surface.

This can then be used to make a flattened representation of the segmented cortical surface. The functional activations elicited by horizontal and vertical meridian stimulation (as estimated in SPM) are then superimposed onto the flattened

representation of the occipital cortex using local software code. As the local representation of the visual field on the cortical surface changes its orientation at the boundary between visual areas, the boundaries can be easily localised (Figure 2.5) and the voxels contained within each visual area exported as a mask image (to be used in later retinotopically specific analyses).

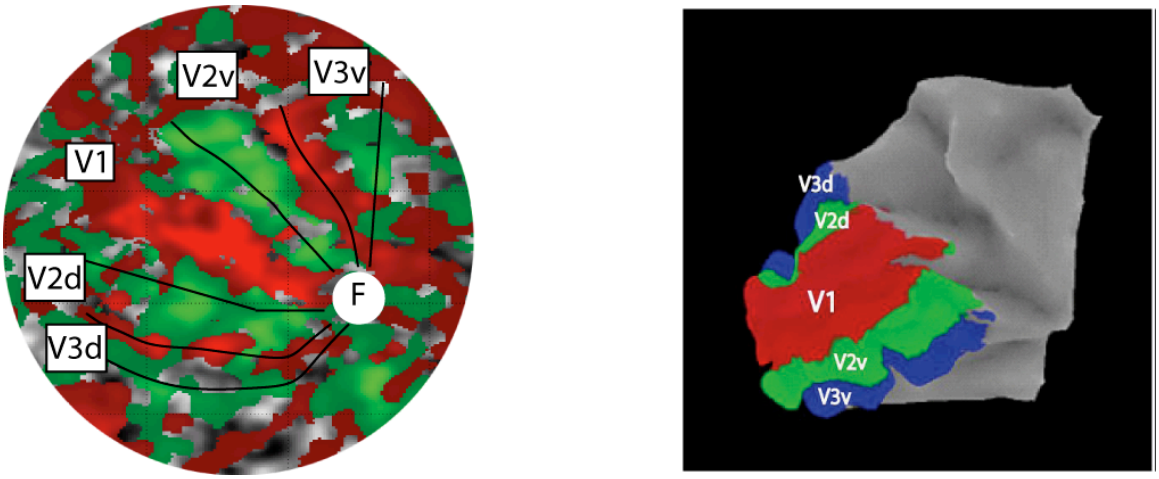


**Figure 2.5. Functional data from meridian mapping projected onto a flatmap of a single subject's left occipital lobe**

The left panel shows functional data from meridian mapping projected onto a flatmap of a single subject's left occipital lobe with boundaries between visual areas added (red represents horizontal – vertical, green represents vertical – horizontal). The right panel shows a 3D reconstruction of the same subject's left occipital lobe with the masks defined in the right panel projected onto its surface.

## 2.5 Conclusion

This chapter has described fMRI and retinotopic mapping, the methods that were used in all of the experiments presented in this thesis. I have presented a summary of the physics and physiology underlying fMRI and the statistical basis of SPM which was used to analyse fMRI data. In addition, I have discussed the physiological basis of retinotopic mapping and how this technique was practically implemented. For practical reasons, the precise use of these methods varied across experiments and each experiment utilised additional methods. Therefore each experimental chapter in this thesis has a methods section describing these points in more detail.

# CHAPTER 3: BRAIN MECHANISMS MEDIATING AUDITORY ATTENTIONAL CAPTURE IN HUMANS.

## 3.1 Introduction

In everyday life, people are often bombarded with different sensory signals, yet can usually focus on stimuli relevant for the task at hand. This can be achieved by using knowledge and expectations to focus attention on task relevant signals rather than competing irrelevant stimuli. Despite this top-down control, a unique stimulus can 'capture attention', even when task-irrelevant. Despite distracting subjects from their current task, such attentional capture may have a survival advantage, as a unique stimulus may often convey important information about the environment. In the visual modality, the effects of attentional capture have been studied using visual search tasks (Theeuwes, 1992;Theeuwes, 1994;Yantis, 1993). When a target in a search display contains an item that is unique on some feature (e.g. a red circle among green circles), this salient feature appears to 'pop out' of the display, making search efficient. If, however, a non-target stimulus has a unique singleton feature, it will typically disrupt search performance. Such interruption of goal-driven attention can be found even when the object is a singleton on a dimension that is never relevant to the task, suggesting that attention was captured by the singleton. This chapter is concerned with investigating brain activity and behavioural effects related to auditory attentional capture.

### *3.1.1 Auditory attentional capture*

In the auditory domain, many previous studies concerning the neural response to deviant auditory stimuli have concentrated on the pre-attentive process of detection of such auditory deviants. Specifically, generation of the electrophysiological potential known as the mismatch negativity (MMN) is associated with the pre-attentive detection of deviant auditory stimuli (Jaaskelainen et al., 2004;Liebenthal et al., 2003;Opitz et al., 1999b;Opitz et al., 1999a;Opitz et al., 2002;Schroger, 1994;Wolff and Schroger, 2001). Most previous studies of the MMN required subjects to passively listen to a stream of auditory stimuli with no measure of the behavioural response to the presence of a deviant auditory stimulus. Without a concurrent

behavioural measure of any distraction, such studies cannot distinguish neural responses associated with acoustic variability per se from those specific to attentional capture. Recent studies have measured the electrophysiological potentials associated with a deviant auditory stimulus and related them to a reaction time measure of behavioural distraction (Berti et al., 2004;Escera et al., 1998;Escera et al., 2001;Rinne et al., 2006;Roeber et al., 2003;Schroger et al., 2000;Schroger and Wolff, 1998). These revealed that the presence of a rare deviant auditory stimulus elicited MMN, N1 and P3a ERP components and slowed reaction times in a subsequent auditory (Alho et al., 1997;Berti and Schroger, 2004;Rinne et al., 2006;Roeber et al., 2003;Schroger et al., 2000;Schroger and Wolff, 1998) or visual task (Escera et al., 1998;Escera et al., 2001). However, the presence of behavioural interference was always associated with increased auditory variability in the stimulus. Thus, such studies cannot distinguish neural responses associated with acoustic variability in the stimulus from those specific to attentional capture.

### *3.1.2 Separating Auditory Attentional Capture from the Mismatch Negativity*

This chapter is concerned with the neural mechanisms associated with behaviourally measured auditory attentional capture using functional MRI in humans. These results were obtained using the behavioural paradigm that established the phenomenon of attentional capture in an auditory search task (Dalton and Lavie, 2004). Subjects searched a sequence of five rapidly presented tones for a target tone that differed in duration from the surrounding non-target tones. Irrelevant variation in the frequency or intensity of one of the non-target tones ("distractor singleton") increased reaction times and error rates suggesting that the irrelevant feature singleton captured attention. However, when the same irrelevant feature singleton was present in the target tone it had no effect on reaction times or error rates (as expected since the target was already a duration singleton). This made it possible to separate cortical responses that were related to increased auditory variation in the stimulus (i.e. the presence of an irrelevant feature singleton regardless of whether it captures attention) from cortical responses that were specific to auditory attentional capture (i.e. those related to presence of a non-target feature singleton that causes behaviourally defined attentional capture).

## 3.2 Methods

### *3.2.1 Subjects*

Twelve young adults (seven females, 18-30 years old, right handed) with normal hearing gave informed consent to participate in the study, which was approved by the joint ethics committee of the Institute of Neurology and National Hospital for Neurology & Neurosurgery.

### *3.2.2 Stimuli*

On each trial, five tones were presented sequentially with an inter-tone interval of 185 ms. Each tone comprised a sine wave of frequency 480 Hz and intensity 90 dB, with a ramp time of 5 ms at each end of the sound wave envelope. The reference for intensity was SPL, measured using a sound meter (Radio Shack 33-2055). One tone was either shorter or longer than the rest and represented the target. The target tone was either 50 ms or 150 ms long while non-target tones were always 100 ms long. The target was always the third or fourth tone in the sequence to enable subjects to hear at least two standard length tones before the target. On two thirds of the trials, additional auditory variation was present on an irrelevant dimension, defined by either frequency or intensity. This variation could be present in either a non-target tone, or a target tone. If present, the frequency singleton had the same duration and intensity as the other tones, but at a frequency of either 440 Hz or 520 Hz. If present, the intensity singleton had identical duration and frequency as the other tones, but intensity was either 70 dB or 100 dB. On distractor singleton trials, the singleton was positioned either directly before or after the target tone. There were thus five types of search trial: singleton absent, distractor singleton present (either frequency or intensity) or target singleton present (either frequency or intensity) (Figure 3.1).

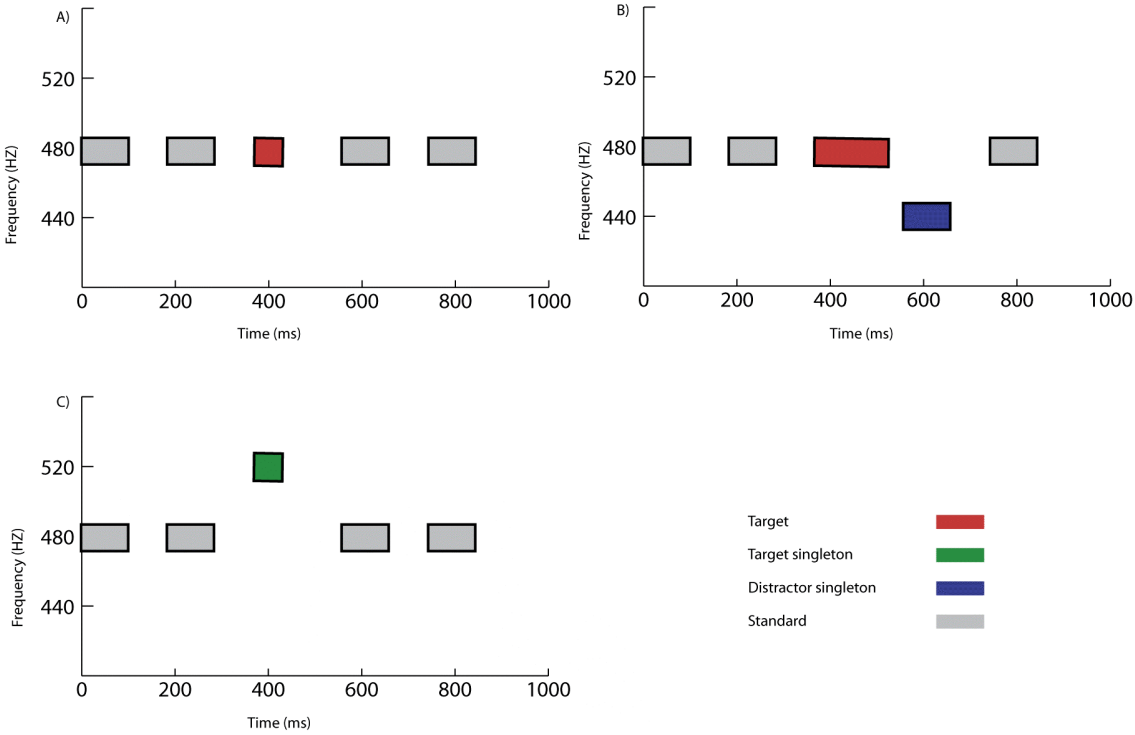


**Figure 3.1 Behavioural conditions**

A. Target only condition. The target was defined on the basis of duration and there was no singleton present. B. Distractor singleton condition. The target was defined on the basis of duration and a distractor singleton was presented with a standard length but a lower frequency. C. Target singleton condition. The target was defined by length and was also presented at a different frequency.

### *3.2.3 Experimental paradigm*

Stimuli were presented binaurally using electrostatic headphones (KOSS, Milwaukee, USA. Model: ESP 950 Medical) custom adapted for use in the scanner. Subjects kept their eyes open and fixated a small cross projected centrally onto a screen mounted approximately 30cm from their eyes and viewed by a mirror mounted on the head coil. Each experimental trial consisted of presentation of the five tones for 925 ms, followed by a 2500 ms response interval. Subjects were required to make a speeded response to the length (short or long) of the acoustic target, by pressing one of two response keys on a keypad held in their right hand. Reaction times (RTs) were measured from the end of the target tone. One quarter of all trials were null trials, on which no sounds were presented. Each participant completed 4 blocks of 128 trials,

divided equally between target singleton, distractor singleton, singleton absent and null trial types. Within separate scanning runs the singleton dimension (intensity or frequency) was kept constant. Trials were pseudo-randomly distributed to optimize the efficiency of estimating the fMRI response. Each participant performed a half-hour practice session prior to entering the scanner in order to ensure they understood and were able to perform the task.

### *3.2.4 Eye position monitoring*

During scanning, eye position was continually sampled at 60Hz using long-range infrared video-oculography (ASL 504LRO Eye Tracking System, Mass). The measures recorded were x and y coordinates of gaze direction (later combined to calculate the distance of eye position from the fixation point), and pupil diameter. Blinks and periods of signal loss were removed from the data, which were subsequently detrended to compensate for eye-tracker drift. The mean position of the eye was computed for each trial and then averaged across trial types within participant.

### *3.2.5 Preprocessing and imaging*

A 3T Siemens Allegra system was used to acquire both T1 weighted anatomical images and T2*-weighted echoplanar (EPI) images with Blood Oxygenation Level Dependent contrast (BOLD). Each EPI image volume comprised of forty 3mm axial slices with an in-plane resolution of 3x3mm positioned to cover the whole brain. Data was acquired in four runs, each consisting of 205 volumes. The first five volumes of each run were discarded to allow for T1 equilibration effects. Volumes were acquired continuously with a TR of 2.6s per volume.

Functional imaging data were analyzed using Statistical Parametric Mapping software (SPM2, Wellcome Department of Imaging Neuroscience, University College London). All image volumes were realigned spatially to the first, and temporally corrected for slice acquisition time (using the middle slice as a reference). Resulting volumes were spatially normalized to a standard EPI template volume based on the MNI reference brain in the space of Talairach and Tournoux (Talairach Tournoux,

1988) and resampled to 2mm isotropic voxels. The normalized image volumes were then smoothed with an isotropic 9mm FWHM Gaussian kernel. These data were analyzed using an event-related random-effects model. Voxels that were activated in the experimental conditions were identified using a statistical model containing regressors that represented the transient responses evoked by the individual trials in each condition. The event-related changes in evoked activity were modeled by convolving an empirically derived hemodynamic impulse response function with trains of unitary events that were aligned on the trial onsets. A trial consisted of a five tone sequence and a response interval. Each component of the model served as a regressor in a multiple regression analysis that included the three experimental conditions and the motion correction parameters (as effects of no interest). The data were high-pass filtered (cut-off frequency 0.0083 Hz) to remove low-frequency signal drifts, and global changes in activity were removed by proportional scaling. The resulting parameter estimates for each regressor at each voxel were then entered into a second level analysis where each participant served as a random effect in a within-subjects ANOVA. Appropriate corrections were made for non-sphericity (Friston et al., 2002) and correlated repeated measures. The main effects and interactions between conditions were then specified by appropriately weighted linear contrasts and determined using the t-statistic on a voxel-by-voxel basis. An initial two factor ANOVA looking at the factors of singleton dimension (frequency or intensity) and singleton type (target singleton or distractor singleton) showed no significant interaction between singleton type and singleton dimension ($p<0.001$uncorrected). In view of this result I subsequently collapsed across the factor of singleton dimension and constructed a one factor ANOVA with three levels of singleton type (target singleton, distractor singleton and singleton absent). A t test was used to identify cortical areas that showed a significant response to auditory attentional capture (distractor singleton > target singleton). Cortical responses to the presence of additional auditory variability were determined by inclusively masking the statistical contrast of distractor singleton versus singleton absent with the statistical contrast of target singleton versus singleton absent ($p<0.0001$ uncorrected). This method was used to isolate cortical areas that were common to both comparisons.
A statistical threshold of $p<0.05_{FDR}$ (Genovese et al., 2002) corrected for multiple comparisons across the entire brain volume was used except for regions that were

hypothesized a priori, where a threshold of $p<0.001$, uncorrected for multiple comparisons was used.

### 3.3 Results

#### *3.3.1 Behaviour*

A two way within subjects ANOVA was initially conducted on the reaction time data. The factors were singleton dimension (frequency or intensity) and singleton type (target singleton, distractor singleton and singleton absent). A significant main effect of singleton type was found [$F(1,11) =15.6$, $p =0.0004$] since reaction times for distractor singleton trials [M RT = 869 ms] were significantly longer than both target singleton trials [M RT = 772 ms; $t(11) = 4.1$; $p = 0.002$] and singleton absent trials [M RT = 749 ms, $t(11) = 4.3$, $p = 0.001$]. There was no difference in reaction times between target singleton and singleton absent trials [$t(11) = 1.5$, $p =0.2$]. The main effect of singleton dimension was not significant [$F(1,11) =0.38$, $p =0.55$]. There was no significant interaction between singleton type and singleton dimension [$F(2,22) = 2.416$, $p = 0.113$].

Error rates were also examined using a similar two way within subjects ANOVA. There was a significant main effect of singleton type [$F(1,11) = 28.5$, $p =0.00001$] since error rates for non target singleton trials [M ER = 18%] were significantly higher than both target singleton trials [M ER = 9%, $t(11) = 5.8$, $p = 0.0001$] and singleton absent trials [M RT = 8%, $t(11) = 7.5$, $p = 0.00001$]. There was no difference in error rates between target singleton and singleton absent trials [$t(11) = 2.0$, $p =0.07$]. The main effect of singleton dimension was not significant [$F(1,11) =0.6$, $p =0.5$]. There was no significant interaction between singleton type and singleton dimension [$F(2,22) = 1.6$, $p = 0.2$]. In addition, behavioural data were analysed to include separate trial types for distractor singleton present before or after the target. There was no significant difference found in reaction times between a singleton presented before the target compared to after the target in both frequency (M RTb = 900 ms, M RTa =868 ms, $t(11) = 1.8$, $p =0.1$) and intensity (M RTb = 873 ms, M RTa =836 ms, $t(11) = 1.1$, $p =0.3$) dimensions. There was also no difference found in error rates when a singleton was presented before the target compared to after the

target in intensity singletons (M RTb = 22%, M RTa =14%, t(11) = 1.8, p =0.06). In the frequency dimension error rates were greater when a singleton is presented before the target rather than after the target (M ERb = 23%, M ERa =13 %; t(11) = 2.89, p =0.02) (Figure 3.2).

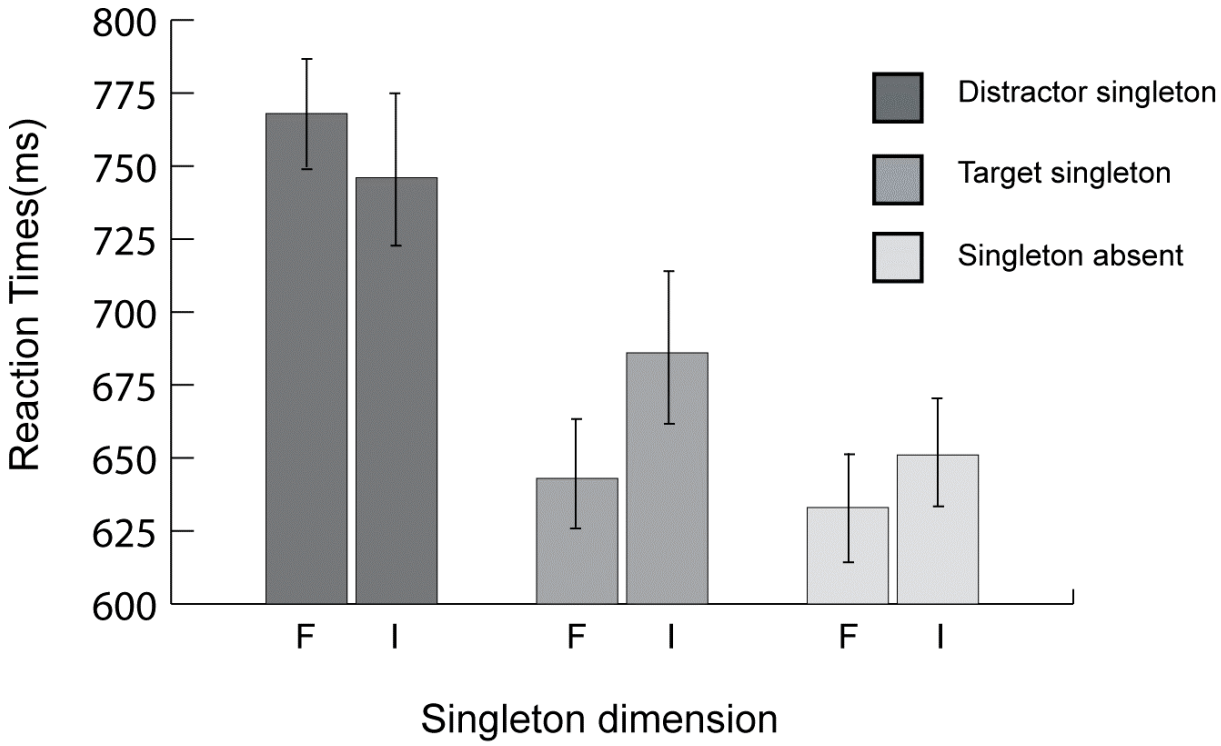


**Figure 3.2 Behavioural results**.

Subjects responded with a button press on each trial to indicate whether the target tone was longer or shorter than the surrounding tones. Mean reaction times averaged across all subjects (n=12) are shown for the three different singleton types: distractor singleton, target singleton and singleton absent for frequency and intensity singletons. The error bars represent the standard error of the mean.

During scanning, eye position was monitored continually with long-range infrared video-oculography in six subjects (see Methods for details). A repeated-measures ANOVA showed no significant differences in mean eye position comparing the different trial types across all subjects ($F(2,10) = 0.631$, $p = 0.463$).

### *3.3.2 Functional MRI*

Preliminary analysis of the functional imaging data confirmed, in agreement with the behavioural findings, that there were no significant ($p < 0.01$, uncorrected) differences between the effects of frequency and intensity singletons on brain activity (see Methods for further details). Therefore for the subsequent analyses I collapsed across frequency and intensity singleton dimension.

#### *3.3.2.1 Singleton presence versus absence*

To identify cortical areas that showed a significant response to singleton presence (irrespective of whether the singleton was present in a target or non-target tone) all singleton present conditions were compared to a singleton absent baseline, using a masking procedure (see Methods). The areas identified by this procedure responded to the presence (versus absence) of a singleton regardless of its type and its behavioural significance, and comprised left inferior frontal gyrus and bilateral superior temporal gyri (Figure 3.3). The stereotactic locations and statistical values for these activated loci are shown in Table 3.1.

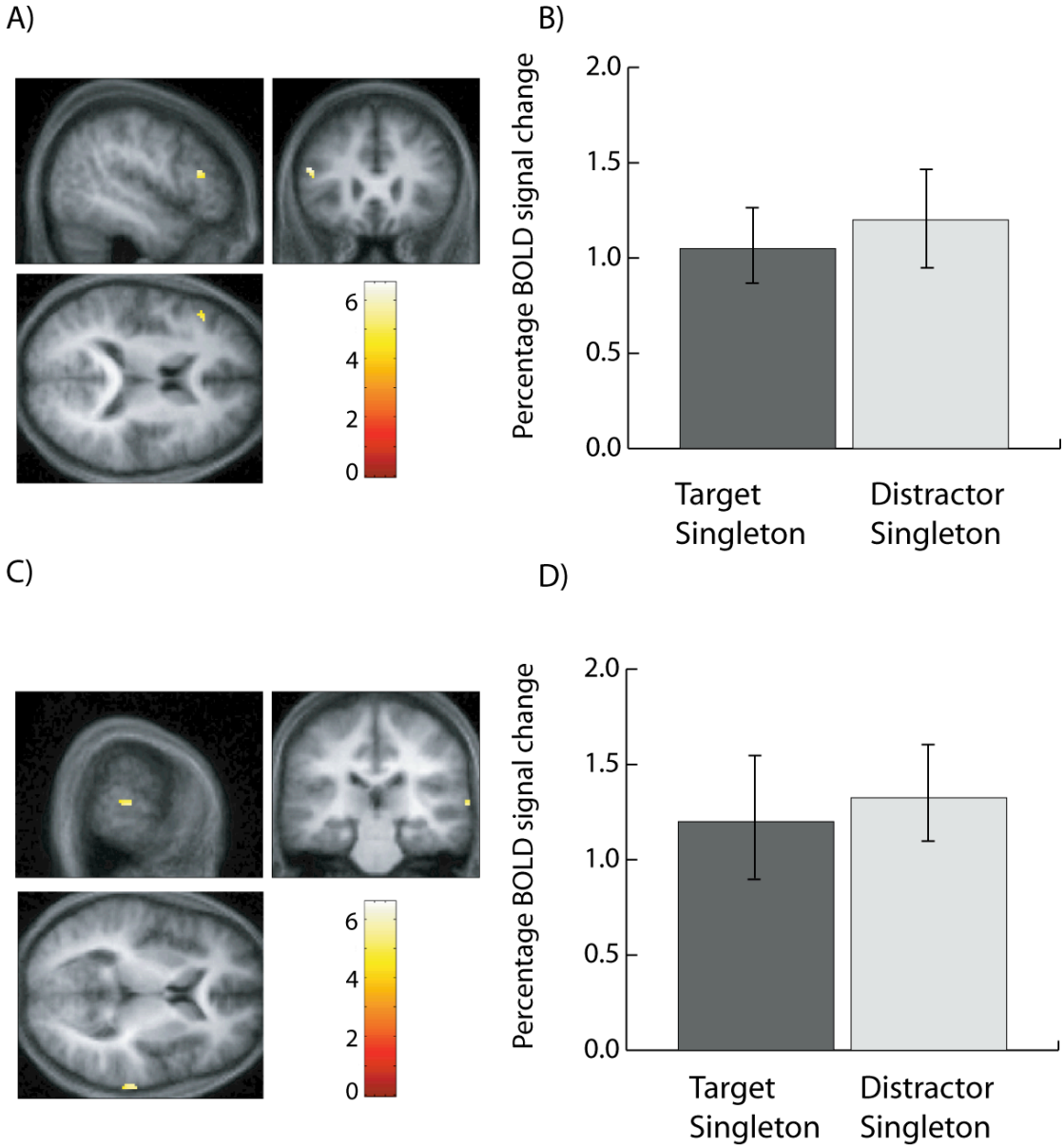


**Figure 3.3 Cortical areas responding to auditory variability.**

Shown in the figure are cortical loci where event-related activity was significantly greater in both distractor singleton trials compared to singleton absent trials, and target singleton trials compared to singleton absent trials ($p<0.05_{FDRcorrected}$). Activated areas are shown projected onto the mean T1-weighted structural scan of the eleven individual subjects. (A) Activated cortical loci in the left inferior frontal gyrus (see Table 1 for stereotactic loci and t values) (B) Percentage BOLD signal in each condition relative to singleton absent base line, averaged across subjects, measured in the left inferior frontal gyrus (C) Activated loci in the right superior temporal gyrus are shown. (D) Percentage BOLD signal change in each condition

relative to a singleton absent base line, averaged across subjects, measured at the right superior temporal gyrus. The error bars in both plots represent the standard error of the mean.

| Anatomy | Coordinates [x y z] | Number of voxels in cluster | t value |
|---|---|---|---|
| L inf frontal gyrus | -44 18 24 | 7 | 6.58 |
| L inf frontal gyrus | -50 26 20 | 28 | 6.47 |
| L sup temporal gyrus | -54, -4, -4 | 15 | 5.86 |
| L sup temporal gyrus | -58 4 -12 | 2 | 4.44 |
| R sup temporal gyrus | 68, -26, 4 | 80 | 6.41 |

**Table 3.1 Coordinates and t values for event-related activation associated with acoustic variability.**

Shown in the table are loci where event-related activity was significantly greater for the comparison of singleton trials (either non-target or target) compared with no singleton trials. Only the most significant peaks within each area of activation are reported in the table ($P<0.05_{FDRcorrected}$).

### *3.3.2.2 Distractor singleton versus target singleton*

The presence of a distractor singleton (compared to target singleton) was associated with activity in a restricted set of parietal and frontal loci: right superior parietal gyrus and intraparietal sulcus and left precentral gyrus (Figure 3.4). The stereotactic loci and corresponding statistical values for each activated locus are shown in Table 3.2.

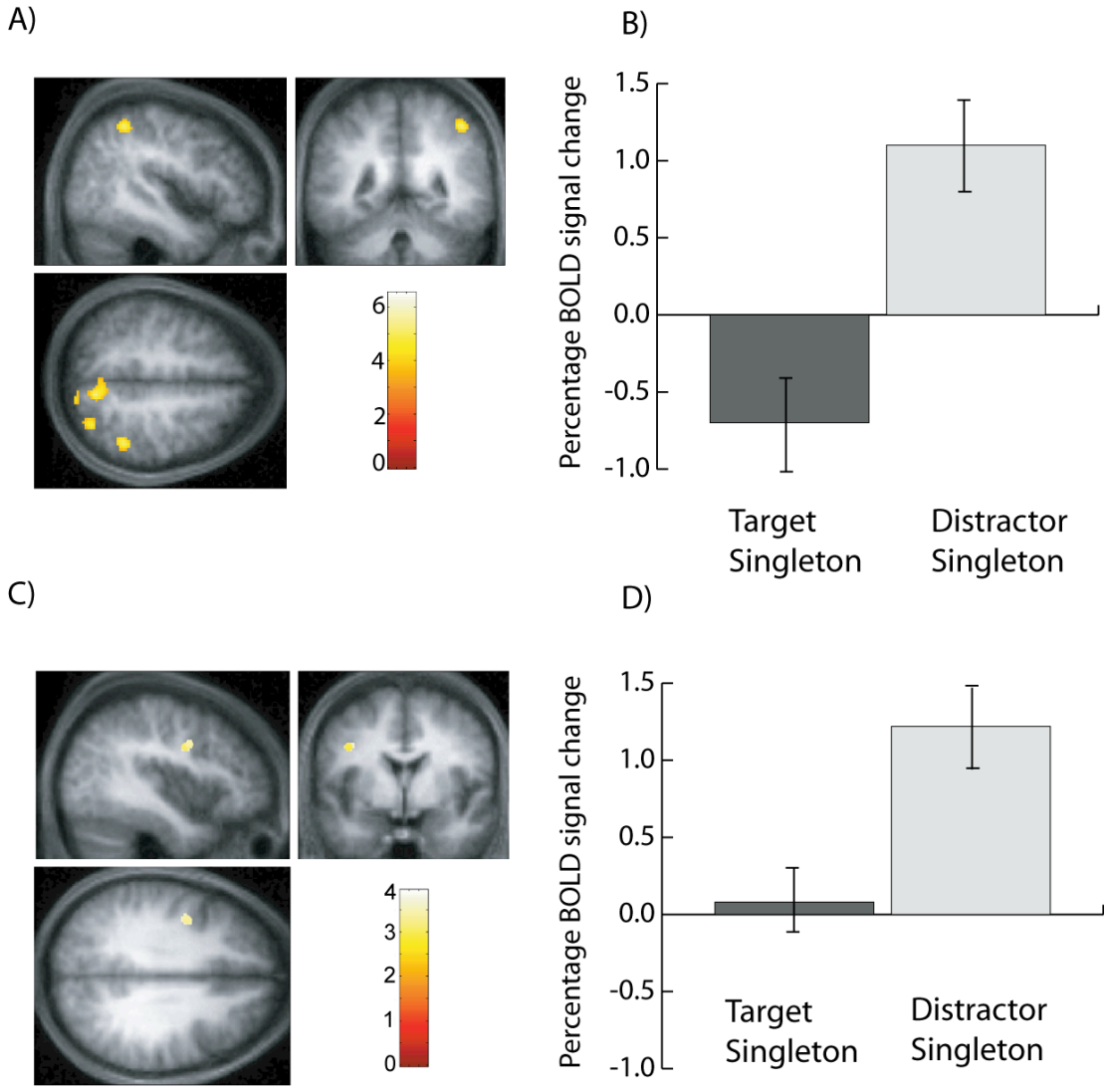


**Figure 3.4.Cortical areas specific for auditory attentional capture.**

Shown in the figure are cortical loci where event-related activity was significantly greater during distractor singleton trials compared with target singleton trials ($P<0.05_{corrected}$). Activated areas are shown projected onto the mean T1-weighted structural scan of the eleven individual subjects. (A) Activated cortical loci in the right superior parietal gyrus and right intraparietal sulcus (see Table 2 for stereotactic coordinates and t values) (B) Percentage BOLD signal in each condition relative to the singleton absent condition, averaged across subjects, measured in the right superior parietal gyrus. (C) Activated cortical loci in the left precentral gyrus (see Table 2 for stereotactic coordinates and t values) projected onto an average T1-weighted structural scan.(D) Percentage BOLD signal in each condition relative to the singleton absent baseline, averaged across subjects, measured at voxel in the left precentral gyrus.

| Location | Coordinates [x y z] | Number of voxels in cluster | t value |
|---|---|---|---|
| R sup parietal gyrus | 10, -68, 56 | 63 | 6.58 |
| R intraparietal sulcus | 46, -46, 52 | 15 | 5.64 |
| R intraparietal sulcus | 32,-70,52 | 2 | 5.35 |
| L precentral gyrus | -40 0 32 | 30 | 3.93 |

**Table 3.2 Coordinates and t values for event-related activation associated with auditory attentional capture.**

Shown in the table are cortical loci where event-related activity was significantly greater during distractor singleton trials compared with target singleton trials ($P<0.05_{FDRcorrected}$ ). Only the most significant peaks within each area of activation are reported in the table.

There was no overlap between the cortical areas that showed activity related to singleton presence (versus absence) compared to those whose activity was specifically related to distractor singleton (versus target) at the corrected threshold. A direct comparison of both contrasts at a liberal uncorrected threshold ($p<0.001$ uncorrected) revealed a small area of overlap in the left inferior frontal gyrus.

Areas that responded specifically to the presence of a distractor singleton were located dorsally while those areas that responded to any stimulus variability were located more ventrally. There were no significant differences in brain activity between non target singletons presented before or after the target (Tmax =4.19 $p>0.6$).

### 3.4 Discussion

This study demonstrates that irrelevant variation in the frequency or intensity of the non-target tones ("distractor singleton") increased reaction times and error rates suggesting that the irrelevant feature singleton captured attention. However, when the same irrelevant feature singleton was present in the target tone it had no significant effect on reaction times or error rates (as expected since the target was already a

duration singleton). This finding is consistent with previous behavioural results in auditory attentional capture (Dalton and Lavie, 2006).

The present findings reveal the neural correlates of behaviourally determined auditory attentional capture by an irrelevant feature singleton in an auditory search task. Importantly, the design of this study allowed me to distinguish between the cortical areas that responded to auditory attentional capture from cortical areas that responded simply to an increase in the variability of the auditory sequence.

### *3.4.1 Auditory odd ball*

Increased auditory variability was associated with enhanced activity in bilateral superior temporal gyri, and left inferior frontal gyrus (Table 3.1 and Figure 3.3), in line with previous findings of neural activity related to auditory change or "odd ball" detection (Doeller et al., 2003;Giard et al., 1990;Molholm et al., 2005;Opitz et al., 1999a;Opitz et al., 2002).

### *3.4.2 Auditory attentional capture*

In contrast, activity specifically related to auditory attentional capture by a distractor singleton (vs. target singleton) comprised a restricted set of cortical areas encompassing right superior parietal gyrus, right intraparietal sulcus and left precentral gyrus (Table 3.2 and Figure 3.4). This activation of fronto-parietal cortex cannot represent a response to stimulus variability per se, as the two singleton conditions (target singleton and distractor singleton) were identical with respect to their acoustic variability. Rather, it is the precise attentional significance (whether the feature singleton was associated with the search target or with a non-target) that determined both whether behavioural interference occurred and whether associated fronto-parietal activation was observed.

### *3.4.3 A crossmodal attentional network*

Voluntary allocation of auditory attention, typically in dichotic listening tasks, is associated in humans with activation in a network of superior frontal and parietal

cortical areas (Benedict et al., 1998;Lipschutz et al., 2002;Pugh et al., 1996;Tzourio et al., 1997;Zatorre et al., 1999). One relevant recent study demonstrates that non-spatial voluntary shifts of attention between auditory and visual streams are associated with activation of right superior parietal lobule, left inferior parietal lobule and right frontal cortex (Shomstein and Yantis, 2004). The present study goes beyond these earlier findings by establishing the cortical areas activated by involuntary shifts of auditory attention associated with capture of attention by an irrelevant feature singleton. The parietal and prefrontal network associated with the voluntary (or here, involuntary) deployment of auditory attention, and with nonspatial shifts of attention between vision and audition (Shomstein and Yantis, 2004) is anatomically very similar to the network of areas proposed to fulfil a similar role in visual attention (for a review see Corbetta and Shulman, 2002). These findings suggest that the frontoparietal network may function as a supramodal attentional system. Consistent with such a hypothesis, our findings in auditory attentional capture resemble those found in associated with stimulus driven visual attentional shifts. In a conceptually similar task, De Fockert and colleagues (2004) found that capture of attention by an irrelevant visual feature singleton (characterized by an odd colour) during performance of a search task based on the stimulus shapes (see Theeuwes, 1992) was associated with activation in left prefrontal and bilateral superior parietal cortices. These loci are very close to those activated in the present study, consistent with a common supramodal network for stimulus driven shifts of attention (Downar et al., 2002).

### *3.4.4 Cortical responses to auditory variability*

In addition to establishing the neural substrates of auditory attentional capture, these results also shed light on how the brain responds to auditory stimulus variability, irrespective of its attentional role. Increased auditory stimulus variability (with the presence of singleton sounds) was associated with activity in left inferior frontal and bilateral superior temporal cortices (Figure 3.3). These areas are similar to the network of areas that are proposed to mediate the generation of the MMN (Deacon et al., 1998;Doeller et al., 2003;Escera et al., 2003;Friedman et al., 2001;Giard et al., 1990;Korzyukov et al., 2003;Molholm et al., 2005;Opitz et al., 1999b;Opitz et al., 1999a;Opitz et al., 2002). Generation of the MMN has been consistently shown to

involve bilateral superior temporal gyri (Deacon et al., 1998;Giard et al., 1990;Opitz et al., 1999b). However the role of the frontal cortex in MMN is less well understood with some studies finding no frontal activation (Opitz et al., 1999a) and others suggesting involvement of the right inferior frontal gyrus (Opitz et al., 2002) or bilateral inferior frontal gyrui (Doeller et al., 2003;Molholm et al., 2005). There is some debate about whether the cranial generators of the MMN vary as a function of the acoustic feature that changes. Several previous studies have suggested that the MMN network may differ slightly depending on the acoustical dimension studied (Molholm et al., 2005;Paavilainen et al., 1991). In contrast some studies find no difference (Sams et al., 1991;Schairer et al., 2001). In this study there was no significant difference in cortical activity between frequency and intensity singleton dimensions. Note, however, that the present experimental paradigm is rather different from those typically used to elicit the MMN. Typically, the MMN is evoked by rare deviant stimuli embedded in very long runs of auditory stimuli. In contrast, the present paradigm involved repeated auditory search in a short run of tones that frequently contained a deviant (either target or distractor singletons). However, the same standard tone throughout out the experimental session therefore allowing subjects to build up a reliable trace of a standard tone against which a deviant tone could be detected. I did not measure electrophysiological brain responses in the present experiment, so cannot know whether MMN occurred to the singletons in the present data. Indeed, some authors (Alho et al., 1997;Liebenthal et al., 2003;Naatanen et al., 1993;Opitz et al., 1999b) have proposed that the MMN is abolished if the deviants are frequent and the runs are short (though see Jaaskelainen and others 2004 for a contrary view). Nevertheless, the striking similarity of the loci activated by singleton presence (versus absence) in the present study with the putative MMN generators is consistent with the notion that this network represents a common cortical mechanism for the detection of acoustic variability (or salience).

### 3.5 Conclusion

Taken together, these new findings suggest that a ventral network, involving bilateral superior temporal gyri and left inferior frontal gyri, responds to auditory variability regardless of its relevance to the behavioural task. In contrast, activation of a more dorsal network comprising of left precentral gyrus, right superior parietal gyrus and

right intraparietal sulcus responds specifically to capture of attention by a feature singleton. The ventral network is anatomically very similar to that previously described for automatic auditory change detection. In contrast, the more dorsal network is closely related to structures activated both in previous studies of visual attentional capture and during voluntary auditory shifts of attention.

In the next chapter I investigate the neural correlates of attentional capture in the visual domain. In particular, I examine behaviour and cortical responses to increasing the salience of the distractor singleton.

# CHAPTER 4: VISUAL ATTENTIONAL CAPTURE

In the previous chapter I examined the cortical networks associated with auditory attentional capture by a task irrelevant distractor (Watkins et al., 2007). In this chapter I now examine the process of attentional capture in the visual domain.

## 4.1 Introduction

A typical visual scene contains far too much information for the human brain to process simultaneously. In order to use visual information to effectively guide behaviour, some kind of selective mechanism is required. Attention is the name given to the process of selecting out the 'important' aspects of a visual scene for further processing, while relegating the rest to limited analysis. The control of visual attention reflects both cognitive ('top down') factors such as knowledge and current goals and stimulus-driven ('bottom up') factors that reflect the salience of sensory information. Typically, these two factors dynamically interact to control where and to what we pay attention (Corbetta and Shulman, 2002).

### *4.1.1 Attentional capture*

In vision, the behavioural effects of attentional capture have been extensively studied, using visual search tasks (Theeuwes, 1992;Theeuwes, 1994;Yantis, 1993). When a target in a search display contains an item that is unique on some feature (e.g. a red circle among green circles), this salient feature appears to 'pop out' of the display, making search efficient. If however a non-target stimulus has a unique singleton feature, it will typically disrupt search performance. Such interruption of goal-driven attention can be found even when the object is a singleton on a dimension that is never relevant to the task, suggesting that attention was captured by the singleton.

Theeuwes has argued that attention is automatically deployed to items in terms of their salience, irrespective of whether these items have features relevant to the task at hand. Subsequently, other researchers have proposed that singleton capture is dependent on the adoption of an attentional set (Folk et al., 1992) or singleton detection mode (Bacon and Egeth, 1994;Yantis and Egeth, 1999), although see

(Theeuwes, 2010;Theeuwes and Burger, 1998) for counter arguments. Whichever is the case, the behavioural data indicate that attention may be captured by a salient stimulus, under the appropriate conditions, even when the event is irrelevant to the current task.

The neural substrates of singleton capture in search have previously been investigated by de Fockert and colleagues (de Fockert et al., 2004). Neural activity was measured via fMRI as subjects viewed similar search displays to Theeuwes (Theeuwes, 1991). Subjects searched for a circle among diamonds, and on 25% of the trials the target was a colour singleton, whereas on another 25% of the trials one of the distractors was a colour singleton. Although there was no measured neural activity specifically related to the colour singleton target, the presence of a colour singleton distractor led to bilateral activation within the dorsal parietal cortex and left frontal cortex relative to when no colour singleton was present. The parietal activity was construed to reflect shifts of attention to the distractor item (Corbetta and Shulman, 2002) and the left prefrontal activity related to top down control processes.

Consistent with a shift of attention to the distractor singleton, Hickey, McDonald and Theeuwes (2006) showed that the presence of a salient distractor modulates the N2pc ERP waveform, which is held to reflect the spatial orienting of attention (Luck and Hillyard, 1994). These data suggest that attention is shifted to a salient distractor before being shifted toward a less salient target.

Although the data of de Fockert et al. (2004) indicate that there is bilateral parietal activity related to the appearance of a singleton distractor, other studies have pointed to a difference between the roles of left and right posterior parietal cortices (PPC) in responding to salient events. In the previous chapter I demonstrated that only the right parietal cortex was activated by auditory attentional capture. There have also been a number of previous TMS studies (utilizing a wide variety of spatial attention tasks) that have shown disruption of spatial attention processes following TMS to the right PPC but not the left (Rushworth and Taylor, 2006). In visual search, Ellison et al. (2003) showed that TMS impaired conjunction search following right parietal TMS. Consistent with the idea that the left and right parietal lobes sub serve different functions is the body of research based on unilateral neglect (Parton et al., 2004). The

syndrome of hemispatial neglect, ipsilateral to the side of the lesion, is more commonly found following right parietal as opposed to left parietal lesions.

Recently, Hodsoll and colleagues have demonstrated that TMS over the right parietal cortex eliminates or reduced the interference effect of the distractor singleton compared to no TMS or TMS over the left parietal cortex. They concluded that the right parietal cortex plays a critical role in attentional capture by a distractor singleton (Hodsoll et al., 2009).

De Fockert proposed that the prefrontal cortex was not involved in the initial transfer of attention to the distractor singleton but instead needed to resolve the competition for selection between the target and the distractor. Supporting this hypothesis they demonstrated a significant negative correlation between the activity measured in the left frontal cortex and the subject's reaction time to find the target.

To investigate these issues further I used fMRI to investigate the cortical mechanisms underlying visual attentional capture by a task irrelevant distractor singleton. I hypothesized that increasing the salience of the distractor singleton would result in increasing behavioural interference with visual search (i.e. longer reaction times to find the target). I also hypothesised that cortical areas responsible for top down modulation of attentional capture would be affected by the salience of the distractor singleton. As the salience of the distractor increased, greater top down control would be required to overcome the distractor and complete the task. However, areas that are critical to salience driven attentional capture should be active as long as the distractor is salient enough to capture attention (i.e the distractor singleton is more salient than the target).

## 4.2 Methods

### *4.2.1 Subjects*

Twelve young adults (four females, 18-30 years old, right handed) with normal vision and normal colour vision gave informed consent to participate in the study, which was

approved by the joint ethics committee of the Institute of Neurology and National Hospital for Neurology & Neurosurgery.

### *4.2.2 Stimuli*

The visual search display consisted of five shapes that were equally spaced in a circular arrangement with a radius of 3.1° from a central fixation point. On each trial, four of the shapes were non target diamonds (size 1.5° diagonally). The target shape was always a circle (diameter 1.5°). In the centre of each shape was a line segment (length 0.5°) randomly chosen to have either a horizontal or vertical orientation. These line segments were always presented in white and the stimuli were presented on a grey background. On two thirds of the trials, one of the shapes was a colour singleton. This colour singleton could be present in either a non-target diamond (distractor singleton), or a target circle (target singleton). There were thus three types of search trial: singleton absent, distractor singleton present or target singleton present. The salience of the colour singleton feature was varied by changing its colour relative to the background non target stimuli. The colour of the singleton varied from bright red to dark green on an isoluminant line through colour space.

### *4.2.3 Procedure*

Stimuli were projected onto a screen approximately 300mm from the participant's eyes and viewed by a mirror mounted on the head coil. Each experimental trial consisted of presentation of the visual search display for 400ms, followed by a 1700ms response interval. Subjects were required to make a speeded response to the orientation (horizontal or vertical) of the line segment in the target circle by pressing one of two response keys on a keypad held in their right hand. One quarter of all trials were null events on which only the fixation point was presented for the duration of the trial. Eleven of the subjects completed eight blocks and one participant completed six blocks of 140 trials. Each participant performed a half-hour practice session prior to entering the scanner in order to ensure they understood and were able to perform the task.

### *4.2.4 Eye position monitoring*

During scanning, eye position was continually sampled at 60Hz using long-range infrared video-oculography (ASL 504LRO Eye Tracking System, Mass). The measures recorded were x and y coordinates of gaze direction (later combined to calculate the distance of eye position from the fixation point), and pupil diameter. Blinks and periods of signal loss were removed from the data, which were subsequently detrended to compensate for eye-tracker drift. The mean eye position was then computed for each trial and an ANOVA used to establish whether any statistically significant differences in eye position occurred in the different experimental conditions.

### *4.2.5 fMRI scanning*

A 3T Siemens Allegra system was used to acquire both T1 weighted anatomical images and T2*-weighted echoplanar (EPI) images with Blood Oxygenation Level Dependent contrast (BOLD). Each EPI image volume comprised of forty 3mm axial slices with an in-plane resolution of 3x3mm positioned to cover the whole brain. Data was acquired in six to eight runs, each consisting of 120 volumes. The first five volumes of each run were discarded to allow for T1 equilibration effects. Volumes were acquired continuously with a TR of 2.6s per volume.

### *4.2.6 Data analysis*

Functional imaging data were analyzed using Statistical Parametric Mapping software (SPM2, Wellcome Department of Imaging Neuroscience, University College London). All image volumes were realigned spatially to the first, and temporally corrected for slice acquisition time (using the middle slice as a reference). Resulting volumes were spatially normalized to a standard EPI template volume based on the MNI reference brain in the space of Talairach and Tournoux and resampled to 2mm isotropic voxels. The normalized image volumes were then smoothed with an isotropic 9mm FWHM Gaussian kernel. These data were analyzed using an event-related random-effects model. Voxels that were activated in the experimental conditions were identified using a statistical model containing regressors that

represented the transient responses evoked by the individual trials in each condition. The event-related changes in evoked activity were modelled by convolving an empirically derived hemodynamic impulse response function with trains of unitary events that were aligned on the trial onsets and lasted for the duration of the trial. Each component of the model served as a regressor in a multiple regression analysis that included the three experimental conditions and the motion correction parameters (as effects of no interest). The data were high-pass filtered (cut-off frequency 0.0078 Hz) to remove low-frequency signal drifts, and global changes in activity were removed by proportional scaling. The resulting parameter estimates for each regressor at each voxel were then entered into a second level analysis where each participant served as a random effect in a within-subjects ANOVA. Appropriate corrections were made for non-sphericity (Friston et al., 2002) and correlated repeated measures. A normalised measure of the mean RT across the group was created in the distractor and target singleton conditions (across the four values of salience). This regressor was used to find cortical areas whose activation significantly followed the subject's reaction time. A t test was used to identify cortical areas that showed a significant response to visual attentional capture (distractor singleton > target singleton). Cortical responses to the level of distractor salience were determined by inclusively masking the statistical contrast of distractor singleton versus target absent with the statistical contrast of the subject's reaction time ($p<0.0001$ uncorrected). A statistical threshold of $p<0.05_{FDR}$ (Genovese et al., 2002) corrected for multiple comparisons across the entire brain volume was used.

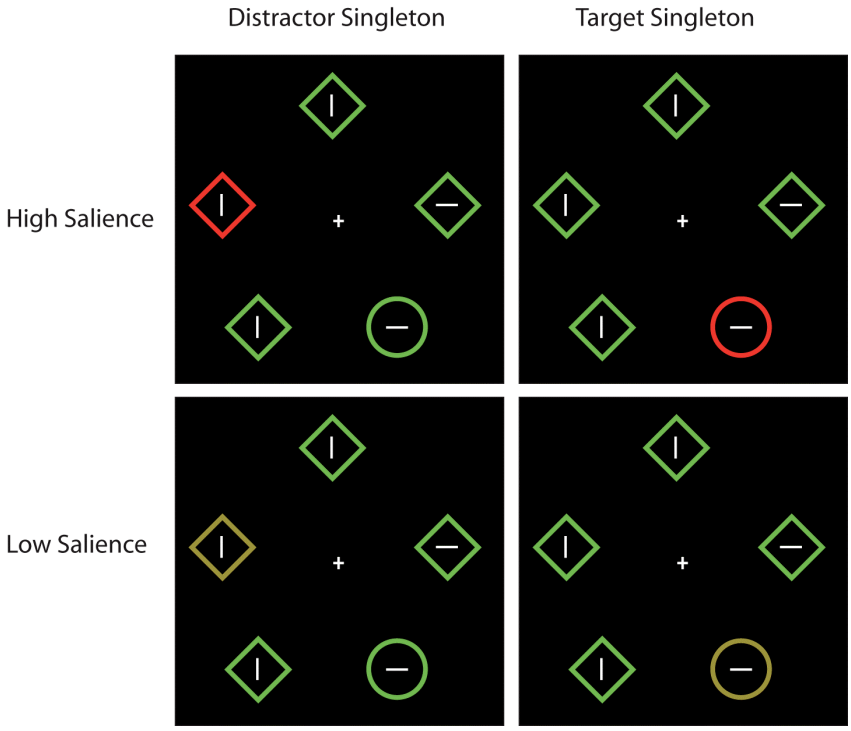


**Figure 4.1 Experimental stimuli.**

Subjects were required to make a speeded key-press response to the orientation of the line segment in the target circle. On singleton present trials, one of the display items (the circle on target singleton trials, and one of the diamonds on distractor singleton trials) was presented in a different colour. The colour of the singleton varied from red (high salience) to dark green (low salience). Note the luminance of the singleton was kept constant.

## 4.3 Results

### *4.3.1 Behaviour*

I initially conducted a one-way within-subjects ANOVA on the reaction time data. The factor was singleton position (target singleton, distractor singleton and singleton absent), collapsed across salience level. A significant main effect of singleton type was found [$F(2,34) = 38.5$, $p<0.0001$] since reaction times for distractor singleton trials [M RT = 759 ms] were significantly longer than both target singleton trials [M RT = 683 ms, $t(11) = 8.2$, $p < 0.001$] and singleton absent trials [M RT = 703 ms, $t(11) = 6.2$, $p < 0.001$]. Target singleton trials [M RT = 683 ms] were significantly shorter than singleton absent trials [M RT = 772 ms, $t(11) = -4.6$, $p = 0.001$].

There was a main effect of singleton level when the singleton was present in the non target (distractor singleton) [F(3,33) = 6.5 , p=0.001] but not in the target singleton [F(3,33) = 0.5, p=0.63].

The error rates were also examined using a similar one way within subjects ANOVA. There was a significant main effect of singleton type [F(2,34) = 7.9,p =0.003] since error rates for distractor singleton trials [M ER = 9.45%] were significantly higher than both target singleton trials [M ER = 6.70%, t(11) = 3.2;p = 0.008] and singleton absent trials [M RT = 6.68%, t(11) = 3.6, p = 0.004]. There was no difference in error rates between target singleton and singleton absent trials [t(11) = 0.43, p =0.97].

There were no statistically significant main effects of singleton salience level on error rates when the singleton was present in the target singleton [F(3,33) = 1.4, p=0.22] or in the distractor singleton [F(3,33) = 3.04, p=0.05].

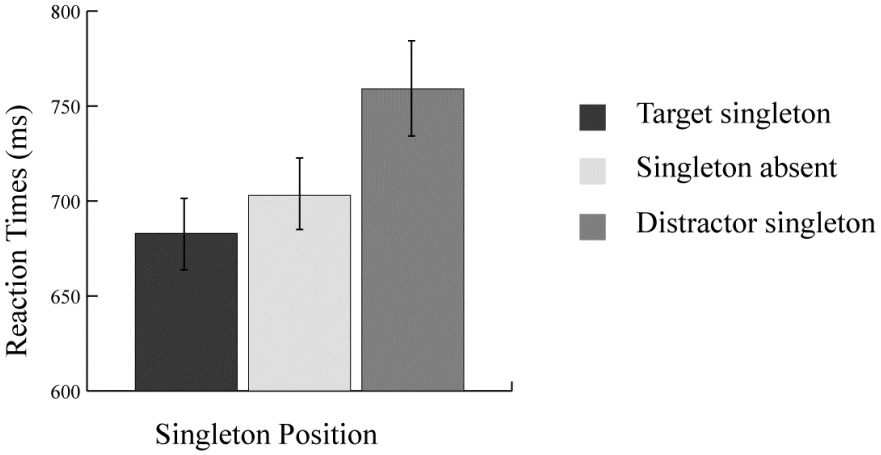


**Figure 4.2 Behavioural results I).**
Mean reaction times averaged across all subjects (n=12) are shown for the three different singleton types: distractor singleton, target singleton and singleton absent averaged across the levels of salience of the singleton. The error bars represent the standard error of the mean.

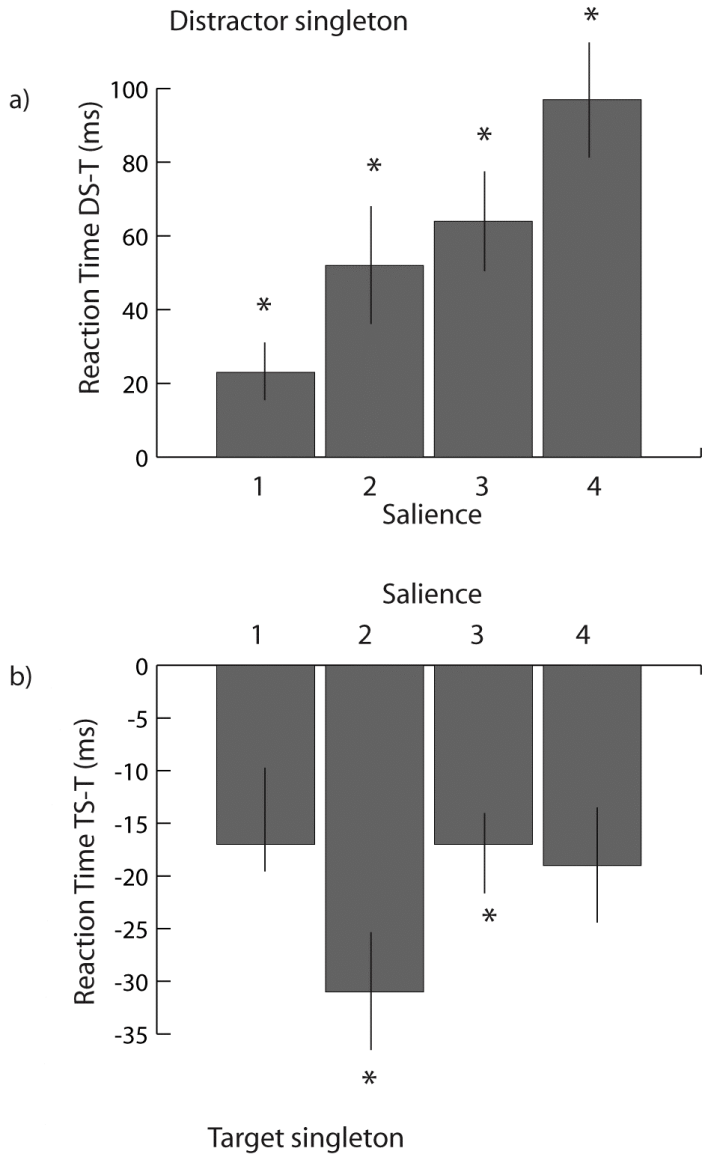


**Figure 4.3 Behavioural results II)**

Mean reaction times averaged across all subjects (n=12) are shown for the four levels of salience (level 4 is the highest salience) in the a) distractor singleton condition and b) target singleton condition. The error bars represent the standard error of the mean. The * represents statistical significance $p<0.05$.

During scanning, eye position was monitored continually with long-range infrared oculography in six subjects (see Methods for details). A repeated-measures ANOVA showed no significant differences in mean eye position comparing the different trial types across all subjects ($F(2,22) = 1.38, p = 0.28$).

### *4.3.2 Functional MRI*

#### *4.3.2.1 Distractor singleton versus target singleton*

The presence of a distractor singleton (compared to target singleton) was associated with activity in bilateral superior parietal and frontal areas (Figure 4.4). The stereotactic loci and corresponding statistical values for each activated locus are shown in Table 4.1.

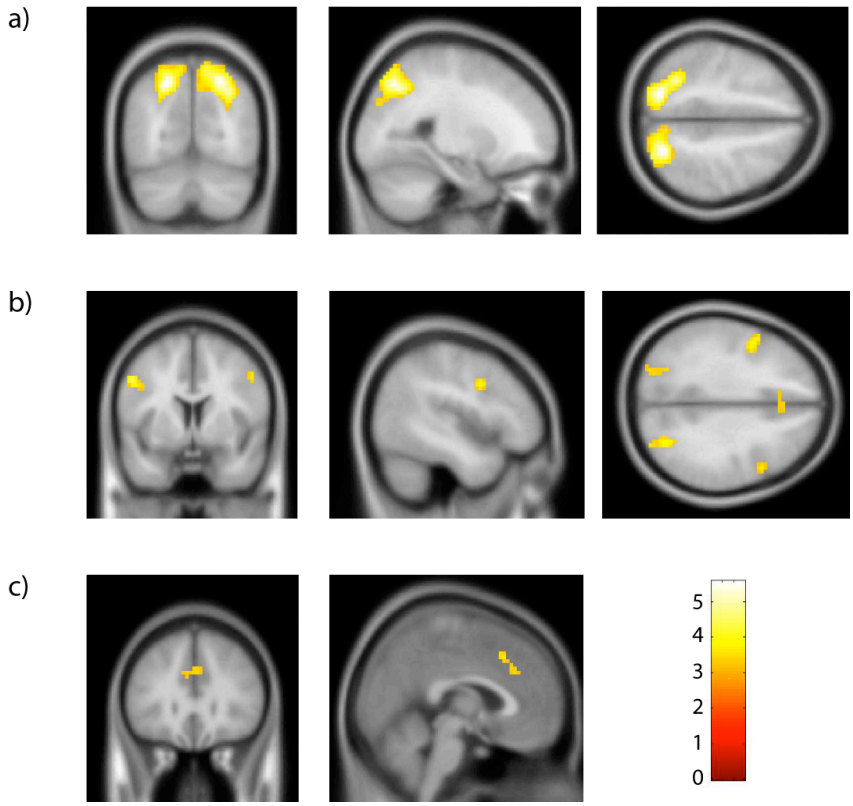


**Figure 4.4. Cortical areas specific for visual attentional capture.**
Shown in the figure are cortical loci where event-related activity was significantly greater during distractor singleton trials compared with target singleton trials ($P<0.05_{FDRcorrected}$). Activated areas are shown projected onto an average T1-weighted structural scan. (a) Activated cortical loci in bilateral superior parietal cortex, (b) Activated cortical loci in the bilateral dorsal frontal cortex and (c) Activated cortical loci in bilateral medial frontal cortex. (see Table 4.1 for stereotactic coordinates and t values)

| Anatomy | Coordinates [x y z] | Number of voxels in cluster | t value |
|---|---|---|---|
| L Superior parietal lobule | -21 -69 48 | 507 | 5.57 |
| R Superior parietal lobule | 24 -69 48 | 477 | 5.54 |
| L Lateral precentral gyrus | -51 6 36 | 48 | 4.45 |
| R Lateral precentral gyrus | 51 9 39 | 28 | 3.77 |
| L Superior frontal gyrus | -27 -3 60 | 10 | 3.7 |
| Bilateral mesial frontal lobe | 6 21 42 | 44 | 3.5 |
| Rt Superior frontal gyrus | 30 0 60 | 10 | 3.5 |

**Table 4.1 Coordinates and t values for event-related activation associated with visual attentional capture.**

Shown in the table are cortical loci where event-related activity was significantly greater during distractor singleton trials compared with target singleton trials ($P<0.05_{FDRcorrected}$ ). Only the most significant peaks within each area of activation are reported in the table.

### *4.3.2.2 Salience*

A relatively restricted set of cortical areas demonstrated a response to the level of salience of the distractor singleton (Figure 4.5). The stereotactic loci and corresponding statistical values for each activated locus are shown in Table 4.2.

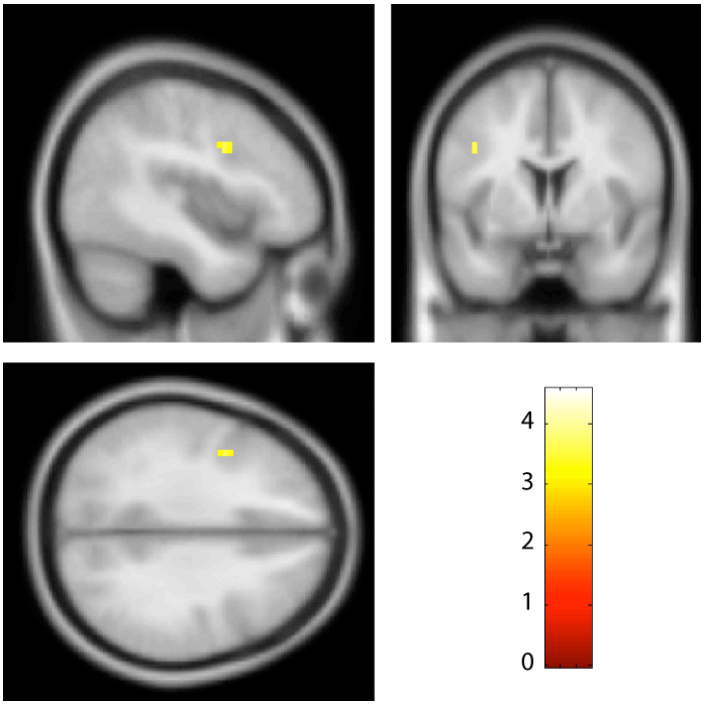

**Figure 4.5 Cortical areas responding to distractor salience.**
Shown in the figure are cortical loci where event-related activity was significantly correlated with the level of distractor salience ($P<0.05_{FDRcorrected}$). Activated areas are shown projected onto the mean T1-weighted structural scan. (a) Activated cortical loci in the left lateral precentral gyrus (see Table 4.2 for stereotactic loci and t values).

| Anatomy | Coordinates [x y z] | Number of voxels in cluster | t value |
|---|---|---|---|
| L Superior parietal cortex | -30 -57 51 | 3 | 4.58 |
| L Lateral precentral gyrus | -42 3 33 | 6 | 3.58 |
| R Lateral precentral gyrus | 48 6 42 | 1 | 3.31 |
| R Superior frontal cortex | 30 -3 60 | 2 | 3.3 |

**Table 4.2 Coordinates and t values for event-related activation associated with increasing distractor salience.**
Shown in the table are cortical loci where event-related activity was significantly greater during distractor singleton trials compared with target singleton trials ($P<0.05_{FDRcorrected}$ ). Only the most significant peaks within each area of activation are reported in the table.

#### *4.3.2.3 Singleton presence versus absence*

There were no cortical areas that demonstrated a response to singleton presence versus absence.

## 4.4 Discussion

This study demonstrated that activation of a bilateral parietal and frontal network was associated with behaviourally defined attentional capture by an irrelevant feature singleton. Importantly, I also demonstrated that activation of bilateral frontal and left parietal areas correlated with the salience of the singleton.

### *4.4.1 Attentional capture.*

The cortical areas associated with attentional capture in this study include a network of bilateral superior parietal cortex, bilateral prefrontal cortex and mesial frontal cortex. They are similar but more extensive to the areas demonstrated previously (de

Fockert 2004). The current experiment has considerable more trials (1120 per subject compared to 480) and subjects (12 versus 10) than that earlier study which provides substantially greater power and may therefore account for our findings of a more extensive cortical network.

### *4.4.2 The role of the dorsal frontoparietal network in attentional capture.*

A dorsal frontoparietal network whose core regions include the dorsal parietal cortex, particularly the intraparietal sulcus / superior parietal lobule and the dorsal frontal cortex along the precentral sulcus, has long been thought to be involved in the top down control of attention. This dorsal system is also thought to generate and maintain endogenous signals based on current goals and sends top-down signals that bias the processing of appropriate stimulus features and locations in sensory cortex. A second ventral network which is strongly lateralized to the right hemisphere, is thought to detect behaviourally relevant stimuli and works as an alerting mechanism or 'circuit breaker' to the first system (Corbetta and Shulman, 2002).

This ventral network was initially considered a good candidate for mediating orientation to salient but unimportant stimuli. However, recent visual studies have shown that activation of the right lateralised ventral network critically depends on the behavioural relevance of the particular object that captures attention (specifically, sharing a feature with the target of the search) (Kincade et al., 2005) This study supports this earlier work by demonstrating that a uninformative but salient distractor activates the dorsal rather than the ventral attentional network.

This study is also consistent with the previous fMRI study on attentional capture confirming bilateral parietal involvement.

I also demonstrated activation in the mesial frontal lobe associated with visual attentional capture. This cortical area has previously been shown to be involved in response error, decision uncertainty and pre response conflict (Ridderinkhof et al., 2004). In the case of this study there were significantly higher error rate in the distractor singleton condition compared to the target singleton condition. Therefore it is possible that the activity shown is related to increased response error.

### *4.4.3 The control of distraction.*

The present theory behind attentional capture suggests that during the first sweep of information through the brain (<150ms) visual selection is completely stimulus driven. Later in time, top down modulation is thought to have an effect in order to resolve response competition and shift attention back to the target (Theeuwes, 2010). Such an account predicts that as the salience of the distractor increases cortical areas responsible for top down control would also increase their activity.

Consistent with such an account, in the current study the bilateral frontal cortices and the left parietal cortex demonstrated increased activity when the salience of the distractor increased. I cannot comment on the time course of this activity because the temporal resolution of fMRI is not fine enough to provide this detail. However my observation of salience-associated activation of these cortical structures is consistent with these areas having a role in the top down modulation of attentional capture.

### *4.4.4. The right superior parietal cortex - a critical area for attentional capture?*

In a recent TMS study by Hodsoll and colleagues (2009), observers performed an attentional capture task while the parietal cortex was stimulated using rTMS. Overall they found a large interference effect when a distractor was present; observers were 137ms slower to find the target than when it was absent. However, the size of the interference effect was reduced to 97ms when rTMS was applied to the right superior parietal lobule (placing the TMS coil according to the stereotactic coordinates from the study of deFockert (2004)) but not when rTMS was applied to the left parietal cortex or when no TMS was applied. In a second experiment the colour singleton was always the distractor (i.e. no target singleton condition). This design reduced the size of the interference effect to ~90ms. Once again rTMS stimulation to the left parietal cortex had no effect on behavioural interference while stimulation of the right parietal cortex caused the interference effect to completely disappear.

In other words when the right parietal cortex is disabled by TMS, a stimulus driven shift of attention to the distractor singleton will not occur. In context of the theory discussed above the right parietal cortex would be postulated to be involved in the

initial shift of attention to a salient stimulus. These findings are generally consistent with previous work suggesting that rTMS over the right PPC disrupts the early phase of the N2pc ERP waveform, which is believed to correspond to attention orientating to a target location in congruent search. This is the same ERP component that was modulated in the presence of a salient distractor in the study of Hickery et al 2006. In the current study we demonstrated strong right PPC activation in response to a distractor singleton versus a target singleton (Figure 4.4). This activity did not significantly increase with distractor salience. This would suggest that the activity in the right PPC may be related to the initial bottom up attentional capture rather than the later top down modulation. (i.e. it always occurs when a singleton is the most salient item in the visual field and attention is captured by it)

### 4.5 Conclusion

I have demonstrated that a bilateral dorsal parietal and prefrontal network of cortical areas were activated by the capture of attention by an irrelevant singleton distractor. Bilateral prefrontal and left parietal cortical activity was correlated with the salience of the singleton suggesting that these areas may be involved in the top-down suppression of the distractor and reorientating attention to the target. The right parietal activity did not increase with the salience of the distractor possibly suggesting in the context of previous work that this cortical area is related to the stimulus driven capture of attention by the distractor.

In the next few chapters I examine the effect of a task irrelevant auditory stimulus on visual perception and cortical processing.

# CHAPTER 5: SOUND ALTERS ACTIVITY IN HUMAN V1 IN ASSOCIATION WITH ILLUSORY VISUAL PERCEPTION.

## *5.1 Introduction*

So far in this thesis I have examined the effects of irrelevant distractors on behaviour and studied the cortical mechanisms associated with attentional capture by a task irrelevant distractor. In the next few chapters I will examine the effect of task irrelevant information on perception. In this chapter I have investigated the effect of a task irrelevant auditory distractor on visual perception. In addition, I have examined the extent to which an auditory induced change in visual perception is represented in primary visual cortex.

### *5.1.1 Audiovisual integration*

The integration of information from multiple senses is fundamental to our perception of the world. Traditionally, it has been assumed that multisensory integration occurs after sensory signals have undergone extensive processing in unisensory cortical regions. However, recent studies in monkey and humans show multisensory convergence at low-level stages of cortical sensory processing previously thought to be exclusively unisensory (for a review see Foxe and Schroeder, 2005). For example, both touch and eye position can influence responses in monkey auditory association cortex and primary auditory cortex, respectively (Fu et al., 2003;Fu et al., 2004;Schroeder et al., 2001). Similarly, primate auditory cortex demonstrates integrated responses to facial and vocal signals of conspecifics (Ghazanfar et al., 2005). These single unit studies are complemented by human event-related potential work demonstrating interactions between auditory and visual (Fort et al., 2002;Giard and Peronnet, 1999;Molholm et al., 2002;Molholm et al., 2004) or somatosensory (Foxe et al., 2002;Murray et al., 2005) stimuli at very short-latency (46 -150ms). One recent fMRI study has shown evidence of audiovisual integration in BA 17 (Calvert et al., 2001) suggesting that primary visual cortex may respond to non-visual inputs. These demonstrations of early modulation of unisensory cortices by multisensory signals challenge hierarchical approaches to sensory processing (Felleman and Van

Essen, 1991;Schroeder and Foxe, 2005) but the function of such multisensory convergence at the anatomically lowest stage of cortical processing remains unclear.

### *5.1.2 Early visual cortex and perception*

An important but unresolved issue that may provide insight into the function of multisensory convergence concerns how such neural interactions might be reflected in conscious perception. If activity in early sensory cortices corresponds to a particular conscious experience, then modification of that activity by converging multisensory input should be related to changes in conscious experience. Behaviourally, the combination of information from different senses can function to reduce perceptual ambiguity (Sumby and Pollack, 1954) and enhance stimulus detection (Bolognini et al., 2005;Frassinetti et al., 2002;Stein et al., 1996). Critically, multisensory convergence can also influence the consciously perceived properties of stimuli (McGurk and MacDonald, 1976;Mottonen et al., 2002;Murray et al., 2004;Murray et al., 2005;Shams et al., 2000;Stein et al., 1996). However, there has been relatively little study of how changes in conscious perception associated with multisensory interactions might be reflected in changes in brain activity.

### *5.1.3 Audiovisual illusion*

In this chapter I have sought to address this issue by measuring brain activity with high field fMRI in human volunteers experiencing an established audio-visual illusion, in which the presence of irrelevant sounds can modify the perception of a simple visual stimulus (Shams et al., 2000). Crucially, this illusion occurs on only a proportion of trials, with veridical perception of the visual stimulus being reported on the non-illusion trials. This means it is possible to compare trials with identical auditory and visual stimulation that nevertheless had very different perceptual outcomes.

Visual evoked potentials and fields are modified at short latency in association with the illusion (Bhattacharya et al., 2002;Shams et al., 2001;Shams et al., 2005), raising the possibility that audio-visual interactions responsible for illusory perception might occur in retinotopic visual cortices. Such a possibility would be consistent with

observations of multisensory interactions in human occipital cortex (Calvert et al., 2001) plus increasing evidence for an association between human V1 activity and unisensory conscious visual experience (Tong, 2003). Therefore I employed retinotopic mapping (Sereno et al., 1995) and specifically focused on activity in retinotopically-defined V1, in order to better study the localisation of any multisensory interactions associated with changes in conscious experience.

## 5.2 Methods

### *5.2.1 Subjects*

Seventeen young adults (8 females, 18-30 years old, right handed) with normal hearing and normal or corrected to normal vision gave written informed consent to participate in the study, which was approved by the local ethics committee. Prior to scanning all subjects took part in a behavioural pilot experiment, following which three subjects were excluded because they did not report the multisensory illusion. Following scanning, two subjects were rejected on the basis of excessive head movement (>5mm) and one subject was rejected because technical problems with the electrostatic headphones. Eleven subjects (8 females, 18-30 years old, right handed) were therefore included in the analysis reported here.

### *5.2.2 Stimuli*

Visual stimuli were projected from an LCD projector (NEC LT158, refresh rate 60 Hz) onto a circular projection screen at the rear of the scanner. The subjects viewed the screen via a mirror positioned within the head coil. The auditory stimuli were presented binaurally using electrostatic headphones (KOSS, Milwaukee, USA. Model: ESP 950 Medical) custom adapted for use in the scanner. All stimuli were presented using MATLAB (Mathworks Inc.) and COGENT 2000 toolbox (www.vislab.ucl.ac.uk/cogent/index.html). Visual stimuli consisted of an annulus with luminance 420cd/$m^2$ and eccentricity 8.5 -10 degrees of visual angle presented for 17ms. The background was a uniform gray screen of luminance 30cd/$m^2$. Luminance calibration was achieved via a viewing aperture in the MRI control room using a Minolta LS-100 spot photometer. An annulus displayed in the peripheral visual field

in association with auditory stimulation was used to maximise illusory perception, which is stronger for stimuli displayed in the periphery (Shams et al., 2002). In addition, the cortical representation of such a peripheral annulus avoids the foveal confluence at the occipital pole (Sereno et al., 1995), where it is extremely difficult to distinguish activity from different early retinotopic visual cortical areas. Our stimulus geometry therefore permitted us to clearly distinguish activity in V1, V2 and V3 from other cortical areas. The auditory stimuli consisted of a sine wave with frequency 3.5kHz, duration 10ms (with a ramp time of 1ms at each end of the sound wave envelope) and volume 95dB. The sound intensity (SPL) produced by the headphones was measured while the headphones were a suitable distance away from the scanner using a sound meter (Radioshack 33-2055).

### *5.2.3 Procedure*

On each experimental trial, subjects were presented with one or two briefly and successively flashed visual stimuli, either alone or accompanied by one or two successively presented auditory bleeps. These comprised six different trial types that represented all the possible combinations of flashes and bleeps. For clarity, these trial types will subsequently be referred to by consistent abbreviations. For example, 'F2B1' refers to trials on which there were two flashes and one bleep while 'F2' refers to a trial on which only two flashes were presented with no auditory stimulation (see Figure 5.1). The interval between flashes in the two flash conditions (F2, F2B1 and F2B2) was 56ms. Pilot behavioural work confirmed that whether bleeps and flashes were presented simultaneously or with slight temporal offset (Shams et al., 2002) made little difference to behavioural reports of illusory perception. On trials with two flashes and one bleep (F2B1), the auditory bleep was presented simultaneously with the first flash. Participants maintained central fixation throughout and indicated whether they perceived one or two flashes, by pressing one of two response keys on a keypad held in their right hand. Each trial lasted 90ms followed by a 1800ms response interval. Eye position data was collected on eight participants during the trials to ensure participants maintained fixation. One eighth of all trials were null trials, during which no visual or auditory stimuli were presented. There were thus seven physically different types of trial. The responses of participants were further used to post hoc divide the F1B2 trials into those on which the illusion was perceived ("F1B2-

Illusion"), and those on which it was not ("F1B2-no Illusion"). Each participant completed between 4 and 8 runs of 128 trials divided equally between the different trial types. Trials were pseudo-randomly distributed within a run.

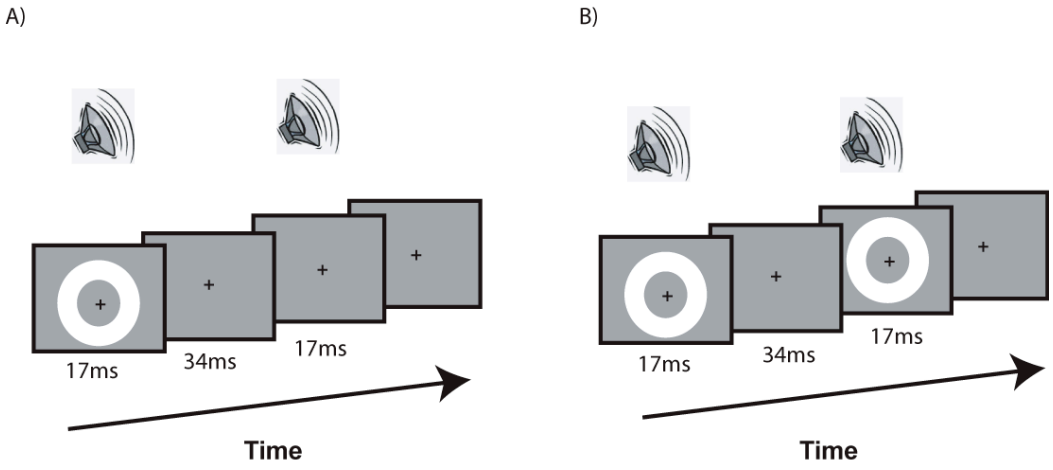


**Figure 5.1 Stimulus configuration.**
Visual stimuli consisted of an annulus presented in the peripheral visual field. The auditory stimuli consisted of a sine wave with frequency 3.5kHz, duration 10ms and volume 95dB. Subjects were presented with one or two briefly and successively flashed visual stimuli, either alone or accompanied by one or two successively presented auditory bleeps. (A) A single visual flash presented with two auditory bleeps (**F1B2**). (B) Two visual flashes presented with two auditory bleeps (**F2B2**). On trials with two flashes and two bleeps (**F2B2**), the bleeps were presented simultaneously with the flashes.

### *5.2.4 fMRI scanning*

A 3T Siemens Allegra system was used to acquire both T2*-weighted echoplanar (EPI) images with Blood Oxygenation Level Dependent contrast (BOLD) and T1 weighted anatomical images. Each EPI image comprised of thirty two 3mm axial slices with an in-plane resolution of 3x3mm positioned to cover the whole brain. Data were acquired in five runs for the first seven subjects, each run consisting of 214 volumes and between six and eight runs, for the last four subjects, each run consisting of 137 volumes. The first five volumes of each run were discarded to allow for T1 equilibration effects. Volumes were acquired continuously with a TR of 2.08s per

volume. During scanning, eye position and pupil diameter were continually sampled at 60Hz using long-range infrared video-oculography (ASL 504LRO Eye Tracking System, Mass). Eye movements were monitored on-line via a video screen for all subjects. Subjects completed a short pilot in the scanner to ensure they could maintain fixation. The eye tracker was calibrated at the start of each experimental run. Eye position was not recorded in three of the subjects for technical reasons.

### *5.2.5 Data analysis*

Eye tracking data were analysed with MATLAB (Mathworks Inc., Sherborn MA). Blinks and periods of signal loss were removed from the data. Mean eye position, expressed as a distance from fixation, was then computed for each trial type and every participant from whom data were available. A repeated-measures ANOVA was used to establish whether mean eye position deviated significantly from fixation, or between conditions.

#### *5.2.5.1 fMRI preprocessing*

Functional imaging data were analyzed using Statistical Parametric Mapping software (SPM2, Wellcome Department of Imaging Neuroscience, University College London). All image volumes were realigned spatially to the first, and temporally corrected for slice acquisition time (using the middle slice as a reference). Resulting image volumes were coregistered to each subject's structural scan. The fMRI data were analyzed using an event-related model. Activated voxels in each experimental condition for each subject were identified using a statistical model containing boxcar waveforms representing each of the experimental conditions, convolved with a canonical hemodynamic response function and mean corrected. Motion parameters defined by the realignment procedure were added to the model as six separate regressors of no interest. Multiple linear regression was then used to generate parameter estimates for each regressor at every voxel for every subject. Data were scaled to the global mean of the time series and high pass filtered (cut-off: 0.0083 Hz) to remove low-frequency signal drifts.

#### *5.2.5.2 Retinotopic analyses*

To identify the boundaries of primary visual cortex, standard retinotopic mapping procedures were used (Sereno et al., 1995;Teo et al., 1997;Wandell et al., 2000). Flashing checkerboard patterns covering either the horizontal or vertical meridian were alternated with rest periods for 16 epochs of 26 s over a scanning run lasting 165 volumes (see Figure 5.2A for details). SPM2 was used to generate activation maps for the horizontal and vertical meridians. Mask volumes for each region of interest (left and right V1, V2d, V2v, V3d, V3v) were obtained by delineating the borders between visual areas using activation patterns from the meridian localisers (see Figure 5.2B for representative meridian maps). The standard definitions of V1 were followed together with segmentation and cortical flattening in MrGray (Teo et al., 1997;Wandell et al., 2000).

Using the mask volumes for left and right V1, V2 and V3, I identified voxels that showed significant activation ($p<.05$ uncorrected) for the comparison of all trials on which visual stimulation was present (i.e. all experimental conditions) compared to null events, employing the regression analysis described above. This comparison identifies voxels activated by the annular visual stimulus in each of the retinotopic areas determined by the independent meridian mapping procedure.

Informal examination of these activations superimposed on flattened representations of occipital cortex confirmed our expectation that they represented voxels activated by our annular visual stimulus (see Figure 5.2B for two representative participants). Having thus independently identified the stimulus representation in V1-V3 the final analytic step was to extract and average regression parameters resulting from the analysis of the main experimental time-series (described above).

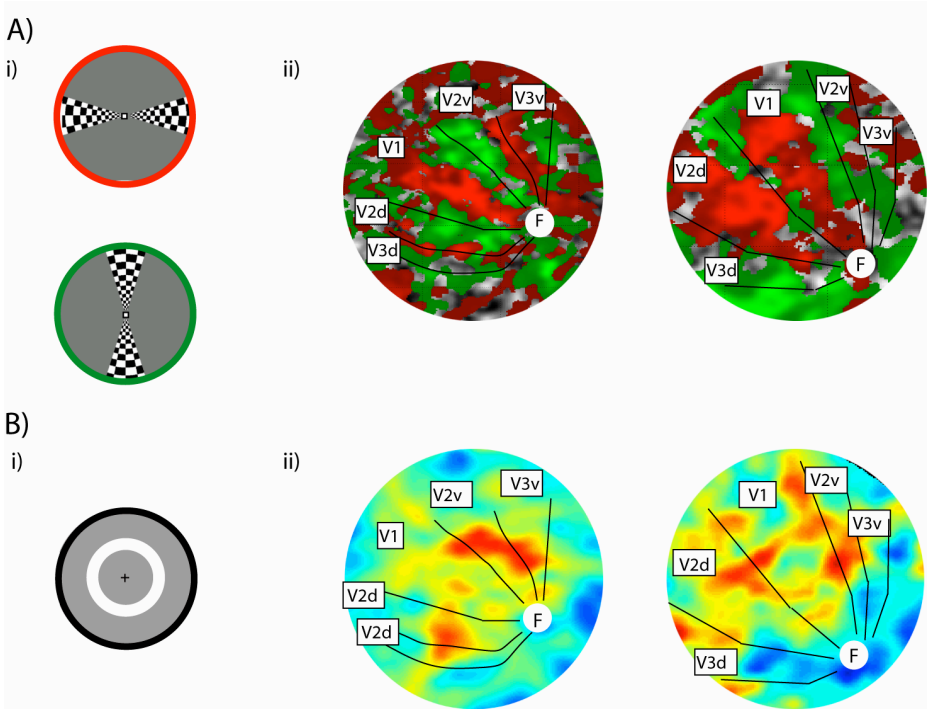


**Figure 5.2 Stimulus representation in visual cortex.**

(A) i) Visual stimuli used to map the horizontal and vertical meridians (see Methods) ii) The outline of individual visual areas V1, V2v, V2d, V3v and V3d and the fovea (determined by meridian mapping, see Methods) are demonstrated for two representative subjects. (B) i) Visual stimulus used in the main experiment ii) The spatial distribution of stimulus-evoked activity (contrast of all visual events (**F1B1**, **F1B2**, **F2B1**, **F2B2**, **F1** and **F2**) versus null events thresholded at $p<.05$ uncorrected) is shown projected onto a flattened representation of visual cortex for two representative subjects (see Methods).

This procedure reliably yielded estimates of percentage signal change for each condition averaged across voxels in V1, V2 and V3 that responded to the visual stimulus for every participant. The percentage signal change was divided by the average cortical response to visual stimulation (i.e. ([F1 +F2]/2) in each subject to produce a normalised percentage signal change for each condition. The statistical significance of any differences in activation between audiovisual and visual trial types was subsequently assessed by entering the normalised percentage signal change for each participant in each of the conditions (F1, F2, F1B1, F2B1, F1B2 no-Illusion, F2B2) into a two-way within subjects ANOVA using a conventional significance level of $p<.05$ (two-tailed). The factors were flash number (1 or 2) and bleep number (0, 1 or 2). The statistical significance of any differences in activation between the

F1B2 Illusion condition and the F1B2-no-Illusion condition was assessed by entering the normalised percentage signal change for each subject in each condition into a two tailed t test using a significance level of $p<.05$ (two-tailed).

Finally, an image of the voxels in V1 that did not show a significant response to the visual stimulus was calculated. This image was then used to repeat the above procedure to examine the response to each condition in the non stimulus responsive area of V1. Corrections for multiple comparisons were made for brain regions where there was no a priori hypotheses. In retinotopic visual cortex I made no correction, as I independently defined the anatomical borders and specific anatomical location activated by the stimuli using orthogonal contrasts; and had specific hypotheses regarding the level of activation in different conditions based on prior work with this paradigm.

#### *5.2.5.3 Whole brain analysis*

To complement the retinotopic analyses, an unbiased examination of regions outside retinotopic cortex was also conducted using a random-effects whole-brain analysis. I did not further examine regions within occipital cortex for this analysis, as it is well established that there is very significant variability in retinotopic areas across individuals (Dougherty et al., 2003). This variability is independent of gyral and sulcal anatomy, it is not taken into account by the spatial normalisation required for this group analysis. The realigned and slice time corrected images from each participant were spatially normalized to a standard EPI template volume based on the MNI reference brain in the space of Talairach and Tournoux (Talairach Tournoux, 1988). The normalized image volumes were then smoothed with an isotropic 9mm FWHM Gaussian kernel. These data were analyzed using an event-related random-effects model, the first stage of which was identical to the regression model described above for the retinotopic analyses, except now applied to spatially normalised images. Activated voxels in each experimental condition for each participant were identified using a statistical model containing boxcar waveforms representing each of the eight experimental conditions, convolved with a canonical hemodynamic response function and mean corrected. Motion parameters defined by the realignment procedure were added to the model as six separate regressors of no interest. Multiple linear regression

was then used to generate parameter estimates for each regressor at every voxel for every participant. Data were scaled to the global mean of the time series and high pass filtered (cut-off: 0.0083 Hz) to remove low-frequency signal drifts. The resulting parameter estimates for each regressor at each voxel were then entered into a second level analysis where each participant served as a random effect in a repeated-measures ANOVA. Appropriate corrections were made for non-sphericity and correlated repeated measures (Friston et al., 2002). The main effects and interactions between conditions were then specified by appropriately weighted linear contrasts and determined on a voxel-by-voxel basis. For these whole brain analyses, a statistical threshold of $p<.05$ for multiple comparisons was used, except for the superior colliculus where a sphere of diameter 4mm centered on the anatomical location of the superior colliculus (as defined by previous studies; (Calvert et al., 2001) was used to apply a small volume correction ($p<.05$, corrected).

## 5.3 Results

### *5.3.1 Behavioural*

Analysis of behavioural responses during scanning confirmed (see Figure 5.3) for a full description of behavioural responses in every condition) that participants were able to accurately report the number of flashes when there was no associated auditory stimulus (i.e. F1 and F2 trial types), when the number of flashes and bleeps were identical (i.e. F1B1 and F2B2 trial types; accuracy 93%, SE across participants 3%) or when two flashes were presented with one bleep (F2B1 accuracy 88% SE across participants 4%). However, on a large proportion of trials when one flash was accompanied by two bleeps (F1B2 trials), participants reported an illusory perception of two flashes ("F1B2-Illusion"; 32% of all F1B2 trials, SE across participants 5%). On the remainder of F1B2 trials, participants reported veridical perception of one flash ("F1B2-no Illusion").

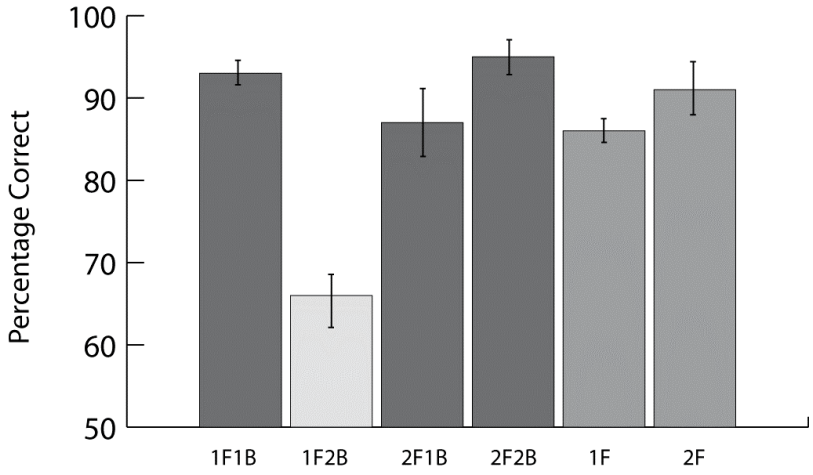


**Figure 5.3 Behavioural results.**

Participants responded with a button press on each trial to indicate whether they saw one or two flashes. Percentage correct responses averaged across all the participants (n=11) are shown for the six different conditions: **F1B1** (one flash with one bleep), **F1B2** (one flash with two bleeps), **F2B1** (two flashes and one bleep), **F2B2** (two flashes and two bleeps), **F1** (one flash presented alone with no auditory stimulation), and **F2** (two flashes presented alone with no auditory stimulation), The error bars represent standard errors of the mean. Note, the low probability of correct responses in the F1B2 condition indicates that on the remaining trials (35%) the participants reported the illusionary perception of two flashes.

Signal detection theory analysis indicated a change in sensitivity (d') between visual stimuli presented alone and visual stimuli presented with concurrent auditory stimuli. Sensitivity d' was defined as d' = z(hits) - z(false alarms), where z is the inverse cumulative normal. For this analysis, double flashes were treated as the target and a correct identification of that stimulus was counted as a 'hit', while the correct identification of a single flash was counted as a 'correct rejection'. 'False alarm', therefore, corresponded to single flash trials on which participants reported seeing two flashes. On average, the presence of two bleeps (d' = 2.67, SD =.47) decreased sensitivity by 15% compared with the visual-alone conditions (d' = 2.28, SD =.60; $t(10) = 2.74$, $p = .02$). The presence of concurrent auditory bleeps was not associated with any significant change in absolute criterion bias ($|\beta| = 1.46$, SD = 2.2) compared to the visual alone condition ($|\beta| = .27$, SD = .2; $t(10) = 1.71$, $p = .12$). Had the illusion been the result of a change in criterion bias, the sensitivity should stay constant; yet I identified significant changes in d' due to the introduction of concurrent auditory

stimuli, suggesting changes in the perceptual processing of the stimulus. This replicates previous findings with this multisensory illusion (Shams et al., 2002) and confirms that illusory multisensory perception can be robustly demonstrated even in the noisy environment of the fMRI scanner.

### *5.3.2 Eye position data*

Participants were requested to maintain fixation at the centre of the display. During scanning eye position was monitored on-line in all participants to ensure participants successfully maintained fixation throughout the experiment sessions. For technical reasons eye data was only recorded in eight participants. A repeated-measures ANOVA showed no statistical difference in mean eye position from fixation, or between conditions for the eight participants in whom eye tracking data were available ($F(7,49) = .485$, $p = .841$).

### *5.3.3 Functional MRI*

#### *5.3.3.1 Retinotopic analyses*

Initially a two-way within subjects ANOVA on the data with correct responses from each visual area was conducted. As our stimulus was circularly symmetric and auditory stimuli presented binaurally measurements were combined across hemispheres to produce a single averaged measure for V1, V2 and V3. The factors were flash number (1 or 2) and bleep number (0, 1 or 2, hereafter referred to as 'B0', 'B1' or 'B2' respectively). I found a significant main effect of flash number (V1:[$F(1,10)=11.8$, $p=.006$] ; V2:[$F(1,10)=9.4$, $p=.01$]; V3:[$F(1,10)=26$, $p=.0004$] ) since responses were larger following two flashes than one. The main effect of bleep number was also significant (V1:[$F(2,20)=10.7$,$p=.006$]; V2:[$F(2,20)=6.9$,$p=.008$]; V3:[$F(2,20)=6.7$,$p=.01$]). This was due to an increased response in retinotopic cortex when a visual event is accompanied by a sound compared to a visual event alone regardless of the number of bleeps (Figure 5.4A and Table 5.11), (B1–B0,[$t(10) = 3.2$;$p =.009$], B2-B1,[$t(10) = 1.25$;$p=.2$] and B2-B0,[$t(10) = 5.1$;$p = .001$). Importantly, there was no interaction of flash number with bleep number (V1:[$F(2,20)=.04$, $p=.9$]; V2:[$F(2,20)=.64$, $p=.5$]; V3:[$F(2,20)=.1$, $p=.8$]). Thus, early

retinotopic visual areas generally showed enhanced activation when a visual stimulus was accompanied by sound, consistent with previous work implicating these structures in multisensory processing (Calvert et al., 2001).

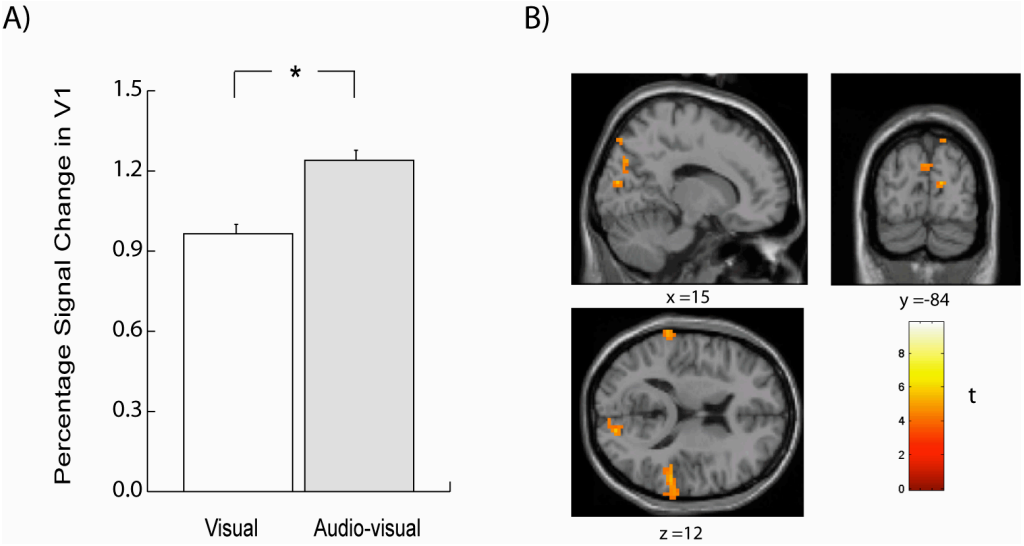


**Figure 5.4 Cortical areas activated by visual stimuli accompanied by auditory stimuli compared to visual stimuli alone.**

(A) Signal change in V1 for visual stimuli alone (**B0**:( **F1** + **F2**)/2) compared with visual stimuli accompanied by sounds (**B1**:(**F1B1** + **F2B1**)/2 and **B2**: (**F1B2-no Illusion** + **F2B2**)/2) collapsed across the number of visual flashes. Data shown are averaged across the eleven subjects (See Methods for further details) with error bars representing the standard error of the mean, and the symbol '*' indicating statistical significance ($p<.05$). (B) Shown in the figure are some of the cortical loci where event-related activity was significantly greater during audiovisual trials compared to visual trials alone (main effect of auditory stimulation; $p<.05$, corrected for multiple comparisons; see also Results). The left hemisphere is presented on the left. The colour of the activation represents the f value, as indicated by the scale bar.

| Location | Coordinates [x y z] | Number of voxels in cluster | f value | p value |
|---|---|---|---|---|
| Rt primary and secondary auditory cortex extending to superior temporal sulcus | 63 -30 9 | 846 | 38.40 | .0001 |
| Lt primary and secondary auditory cortex extending to superior temporal sulcus | -57 -24 3 | 645 | 30.50 | .0001 |
| Lt middle frontal gyrus | -36 -3 36 | 6 | 10.56 | .010 |
| Rt insular | 45 9 9 | 1 | 9.24 | .021 |
| Lt Cerebellum | -24 -69 -24 | 4 | 8.32 | .035 |
| Lt insular | -36 -15 6 | 1 | 8.27 | .036 |
| Anterior cingulate | -15 30 30 | 3 | 8.02 | .042 |
| Rt occipital cortex | 6 -84 30 | 1 | 7.80 | .048 |

Table 5.1 Coordinates and f values for event-related activation associated with the main effect of auditory stimulation ($p<.05_{FDRcorrected}$). Only the most significant peaks within each area of activation are reported in the table.

These findings provide some preliminary evidence regarding multisensory auditory-visual interactions in retinotopic visual cortex. However, our primary goal was to identify correlates of changes in *conscious perception* associated with multisensory interactions. On F1B2 trials, a significant proportion evoked the illusion of two flashes (F1B2-Illusion), while on the remainder only one flash was perceived (F1B2-no Illusion). Next I compared activity in retinotopic visual areas that was evoked on F1B2-Illusion trials with F1B2-no Illusion trials. Note that because I compared physically identical trials with exactly the same visual and auditory stimulation, any differences in brain activity associated with this comparison *cannot* reflect differences in visual or auditory stimulation. I found that activity in V1 was significantly higher for F1B2-Illusion trials on which the illusion was perceived compared to F1B2-no

Illusion when the illusion was not perceived ([t(10) = 2.25, p =.047], two-tailed) (see Figure 5.4 for full details).

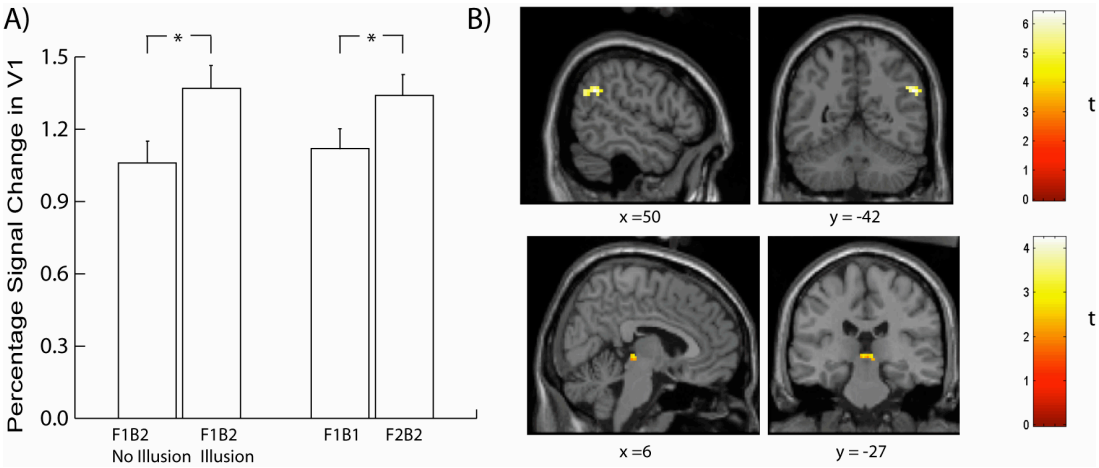


**Figure 5.5. Cortical areas activated by multisensory illusory perception.**
(A) Signal change in primary visual cortex associated with illusory multisensory perception. The mean percentage signal change in retinotopically defined V1 (see Methods) is shown for the condition **F1B2-no Illusion** (one flash with two bleeps when subjects reported correctly the perception of one flash), **F1B2-Illusion** (one flash with two bleeps when subjects reported the illusory perception of two flashes), **F1B1** and **F2B2**. Data shown are averaged across the eleven subjects (see Methods for further details) with error bars representing the standard error of the mean, and the symbol '*' indicating statistical significance ($p<.05$). (B) Shown in the figure are cortical loci outside retinotopic cortex where event-related activity was also significantly greater during **F1B2-Illusion** trials compared to **F1B2-no Illusion** trials ($p<.05$, corrected for multiple comparisons; see also Results). Activated cortical loci in the right superior temporal sulcus (ascending posterior segment) projected onto a sagittal and coronal slice of a canonical T1 template image in the stereotactic space of Talairach & Tournoux (1988). (C) Activated cortical loci in the superior colliculus projected onto a sagittal and coronal slice of the canonical T1 template image. The left hemisphere is presented on the left. The colour of the activation represents the t value, as indicated by the scale bar.

The activity in V1 in the F1B2-Illusion condition was not significantly different to the F2B2 condition [t(10) =.209; p =0.84]. This enhanced cortical response to the illusory perception was specific to the retinotopic locations of V1 responding to the visual annulus, as there was no significant effect of the illusion in the regions of V1 that did not respond to the visual annulus [t(10) = .54,p =.61]. Thus, illusory visual perception

induced by sound is associated with significantly greater activation of stimulus-responsive retinotopic regions of the first visual cortical area, V1. The effect of the auditory-visual illusion on V1 activity was highly consistent across participants, with ten of our eleven participants showing an increase in activity for F1B2-Illusion versus F1B2-no Illusion trials (Figure 5.5b). To visually assess the time course of the event related activations I plotted peristimulus time histograms for the two principal comparisons of interest (F1B2-Illusion and F1B2-no Illusion) averaged across subjects (Figure 5.5a). There was no correlation between the magnitude of the cortical response on F1B2-Illusion trials and the proportion of trials on which each participant experienced the illusion (correlation coefficient = -.148; r squared = .022, p> .1).

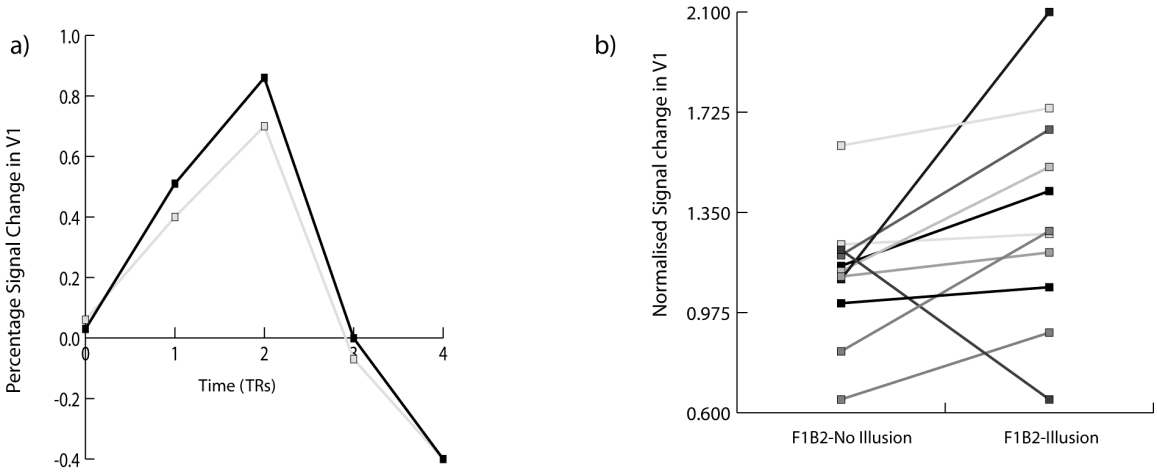


**Figure 5.6. Time courses for V1 cortical responses.**

a) Time courses for the V1 cortical responses in the **F1B2-Illusion** (black line) and **F1B2-no Illusion** (grey line) condition. Percentage signal change in V1 is plotted against time from stimulus onset (units of TR = 2.08 seconds) for both conditions averaged across subjects. The time courses were calculated for each of the subjects by using a statistical model containing a boxcar waveform representing each of the experimental conditions, convolved with a series of FIR (finite impulse response) functions. Motion parameters defined by the realignment procedure were added to the model as six separate regressors of no interest. Multiple linear regression was then used to generate parameter estimates for each regressor at each time point for every subject. The data used in this model was extracted from the area of V1 that responded to the visual stimulus. This was calculated by masking V1 (determined by retinotopic mapping (see Methods)) with the cortical area that showed a significant response ($p<.05$ uncorrected) to of the contrast of all visual events (**F1B1**, **F1B2**, **F2B1**, **F2B2**, **F1** and

**F2**) >**null**. b) Signal change in V1 for the conditions **F1B2-Illusion** and **F1B2-no Illusion** for every subject.

### *5.3.3.2 Whole brain analyses*

To complement the retinotopic analyses, a whole-brain analysis of activity for each of the main comparisons outlined above was also performed (see Methods). A two factor ANOVA was initially conducted. The factors were flash number (1 or 2) and bleep number (0, 1 or 2). As expected, there was a significant main effect of flash number in the left and right occipital cortex (though as this is group data there is no information on which retinotopic area or areas this might represent). The main effect of bleep number was also significant (see Table 1 for a complete listing of these loci and their stereotactic locations, plus Figure 5.3B). This was due to an increased cortical response when a visual event is accompanied by a sound compared to a visual event alone regardless of the number of bleeps [B1–B0,tmax = 7.41;p <.0001, B2-B1,tmax = 4.83;p=.1 and B2-B0,tmax = 8.68;p < .0001]. There was no significant interaction between the number of flashes and the number of bleeps ($t<3.86$, $p > 0.7$). Activation of similar loci have been associated with audiovisual integration (Calvert et al., 2001).

Finally, unrestricted whole-brain analysis of illusory multisensory perception (i.e. F1B2-Illusion vs F1B2-no Illusion) additionally revealed significant activation in the right superior temporal sulcus (ascending posterior segment of the STS, abutting the supramarginal gyrus; co-ordinates [54 -54 30] t=6.83; p = .02 corrected, number of voxels in the cluster = 88) and the superior colliculus (coordinates [2 -30 0], t=3.12 p =.03 corrected for anatomically specified small volume examined; see Methods). These activated loci are shown in Figure 5.4B & 5.4C. There were no cortical areas that showed a significant response to F1B2-no Illusion >F1B2 Illusion.

In summary, the presence of auditory stimuli enhanced visual responses in V1, V2 and V3 (Figure 5.3 and Table 5.1). In contrast, the more restricted comparison of those multisensory F1B2 trials that either evoked the illusion ("F1B2-Illusion") or did not ("F1B2-no Illusion") revealed enhanced activity in primary visual cortex, the superior temporal sulcus and superior colliculus (Figure 5.5).

## 5.4 Discussion

Irrespective of perception, the concurrent auditory stimulation enhanced activity in human V1, V2 and V3. These findings are broadly consistent with previous observations that behavioural or physiological responses to visual stimulation can be modified by sound (Bermant and Welch, 1976;Kitagawa and Ichihara, 2002;Morrell, 1972;Reisberg, 1978;Sekuler et al., 1997). In humans, several functional imaging studies have demonstrated that multisensory interactions can occur in extrastriate regions of visual cortex previously held to be unisensory (Amedi et al., 2001;Calvert et al., 1997;Calvert and Thesen, 2004;Lloyd et al., 2003;Macaluso et al., 2000) and even in Brodmann area 17 of occipital cortex (Calvert et al., 2001). None of these earlier human studies used retinotopic mapping to functionally identify early retinotopic visual cortex. My study therefore extends this important earlier work by explicitly quantifying multisensory effects in retinotopically defined early visual cortex, and confirms that multisensory convergence can occur at the first cortical stages of human visual processing.

Importantly, I further examined how neural interactions associated with multisensory processing might be reflected in conscious perception. Participants were asked to report their perception of the visual stimulus on a trial-by-trial basis, which allowed us to confirm illusory perception of two flashes on a proportion of the F1B2 trials (F1B2-Illusion), as previously reported (Shams et al., 2000). Critically, the brain activity evoked in human V1 on illusion trials (F1B2-Illusion) was significantly greater than on physically identical trials where no illusion was reported (F1B2-no Illusion). Indeed, the cortical activity evoked on F1B2-Illusion trials was not reliably different from F2B2 trials (see Figure 5.4). This effect was robust across participants (Figure 5.5b) and also associated with enhanced activity in the superior colliculus and superior temporal gyrus (Figure 5.4). This enhancement of activity in association with illusory perception did not reflect differences in eye position or eye movements on different trials. Nor did it reflect a general alerting effect caused by changes in auditory stimulation, as both visual and auditory stimulation were identical on illusion and no-illusion trials. Nor was it explained by any bias for our participants to respond incorrectly to the number of bleeps (signal detection analysis revealed a significant change in sensitivity when visual events were accompanied by two bleeps with no

significant change in absolute criterion bias). Instead, the activity in primary visual cortex corresponded better to participants' subjective reports of their visual experience on F1B2 trials, rather than with the physical stimulation (which remained unchanged). That V1 activity can be more closely related to conscious visual experience rather than physical stimulation is increasingly recognised in unisensory studies. For example, activity evoked in human V1 by a visual stimulus briefly presented at the contrast detection threshold is higher on trials when participants successfully detect it than when they fail to do so(Ress and Heeger, 2003). Moreover, when participants falsely perceive the presence of a low-contrast stimulus on trials when the stimulus was physically absent (false alarms), V1 activity is similar to that on trials where participants correctly report the physical presence of a stimulus(Ress and Heeger, 2003). This suggests that for visual stimulation alone, V1 activity can more closely correspond to the contents of consciousness than to visual stimulation. The present findings show that such an association of V1 activity with conscious perception extends to suprathreshold visual stimuli under normal viewing conditions, and to changes in visual perception brought about by multisensory stimulation.

The temporal resolution of fMRI is relatively low, so this study does not provide useful information about the timing of the multisensory effects on visual perception. Rather, it provides precise localisation of these effects to the anatomically lowest stage of cortical processing. Such findings are consistent with the work reviewed earlier suggesting that for sensory modalities such as audition, multisensory influences also extend to the earliest stages of cortical processing (see Introduction and Schroeder & Foxe 2005). Event-related potential recordings in human show that multisensory integration can occur very early in visual processing. For example, a change in a simple visual stimulus that is accompanied by a change in pitch of a concurrent tone can lead to modification of the ERP at very short latencies (Giard and Peronnet, 1999), and auditory clicks can modify the evoked potential to pattern stimulation in visual cortex (Arden et al., 2003). Similarly, a surprisingly early right parieto-occipital interaction between auditory and visual stimulation is seen in the ERP waveform during a simple reaction time task (Molholm et al., 2002). For the particular multisensory illusion reported here, visual evoked potentials and fields associated with the illusory perception are modified at a short latency (Bhattacharya et al., 2002;Shams et al., 2001;Shams et al., 2005) consistent with generators in early

visual cortex. My new findings extend this earlier work by demonstrating that non-visual stimuli can affect early visual processing in human primary visual cortex.

These physiological data do not precisely define how these effects arise, particularly regarding the association of V1 activation with illusory visual perception in the illusion studied here. Primary visual cortex receives projections from at least twelve areas described as belonging to the visual cortex (Felleman and Van Essen, 1991). Recently slightly weaker more distant projections have been described from areas in the ventral (Distler et al., 1993a) and dorsal visual pathways and from the lateral intraparietal area (Boussaoud et al., 1990;Rockland and Van Hoesen, 1994). Several recent papers have used retrograde tracer injections to demonstrate projections from primary auditory cortex, auditory association areas and the superior temporal polysensory area (STP) to the area of primary visual cortex representing the peripheral visual field (Clavagnier et al., 2004;Falchier et al., 2002;Rockland and Ojima, 2003). These connections have a laminar signature and a termination pattern consistent with feedback or lateral type connection(Rockland and Ojima, 2003). The function of these projections to V1 has been the subject of much debate but they may serve to enhance perceptual capabilities; for example the addition of an auditory signal to a visual signal leads to improved detection compared to a visual signal alone (Bolognini et al., 2005;Frassinetti et al., 2002;Gondan et al., 2005;Miller, 1982;Molholm et al., 2002;Schroger and Widmann, 1998c) . Thus, it is possible that these connections mediate the increased activity in V1. Intriguingly, there was a greater cortical response for F1B2 Illusion> F1B2 no-Illusion only in the stimulus responsive area of primary visual cortex. This implies that an auditory stimulus is only effective at exciting primary visual cortex when preceded by a visual stimulus. This finding is consistent with the temporal properties of the illusion (Shams et al, 2000).

When participants experienced illusory visual perception, enhanced activity was not only identified in V1 but also in the superior colliculus and the right superior temporal sulcus (STS), which may also play a role. In particular, the STS may be the human homologue of macaque area STP, and has been consistently associated with integration between visual and auditory stimulation (Beauchamp et al., 2004;Beauchamp, 2005;Calvert et al., 2000;Olson et al., 2002), particularly in speech

recognition (Pekkola et al., 2005;Raij et al., 2000;Saito et al., 2005;Sekiyama et al., 2003;von Kriegstein et al., 2005). A recent study also showed that the STS is involved in crossmodal associative learning (Tanabe et al, 2005). Similarly the superior colliculus is known to play an important role in multisensory integration. Many studies in animals demonstrate the presence of multisensory neurons in the superior colliculus(Meredith et al., 1987;Meredith and Stein, 1983;Stein et al., 1975;Wallace et al., 1998;Wallace and Stein, 1996). Consistent with this, in humans there is an enhanced superior colliculus response to a multisensory signal compared to a unisensory signal (Calvert et al., 2001). Future studies should therefore attempt to elucidate the possible functional role of each structure in the illusion studied here.

### *5.5 Conclusion*

Taken together, the response of retinotopic visual areas V1-V3 to visual stimulation are significantly enhanced by concurrent auditory stimulation. Specifically, when this auditory stimulation gave rise to an illusory change in perceptual experience, this was associated with specific enhancement of in primary visual cortex, superior colliculus and right superior temporal sulcus.

# CHAPTER 6: ACTIVITY IN HUMAN V1 FOLLOWS MULTISENSORY PERCEPTION.

In the previous chapter I demonstrated that when a single brief visual flash is accompanied by two auditory bleeps, it is frequently perceived incorrectly as two flashes. Such illusory multisensory perception is associated with increased activation of retinotopic human primary visual cortex (V1) suggesting that such activity reflects subjective perception (Watkins et al., 2006). However, an alternate possibility is that increased V1 activity reflects either fluctuating attention or auditory-visual perceptual matching on illusion trials. In order to rule out these possibilities I now study the complementary illusion, where a double flash is accompanied by a single bleep and perceived incorrectly as a single flash. In this chapter I replicate findings of increased activity in retinotopic V1 when a single flash is perceived incorrectly as two flashes, and now show that activity is decreased in retinotopic V1 when a double flash is perceived incorrectly as a single flash. These findings provide strong support for the notion that human V1 activity reflects subjective perception in these multisensory illusions.

## 6.1 Introduction

In everyday life our perception of the world is dominated by multi-sensory information. Multi-sensory convergence can influence not only cortical sensory processing (for a review see Foxe and Schroeder, 2005) but also the consciously perceived properties of stimuli (McGurk and MacDonald, 1976;Mottonen et al., 2002;Murray et al., 2004;Murray et al., 2005;Shams et al., 2000;Stein et al., 1996). However, there has been relatively little study of how changes in conscious perception associated with multi-sensory interactions might be reflected in changes in brain activity. In the previous chapter, I used high field fMRI to study brain activity associated with an established audiovisual illusion. When a single brief visual flash is accompanied by two auditory bleeps, it is frequently perceived incorrectly as two flashes (Shams et al., 2000). The perception of this 'fission' illusion is associated with increased activity in retinotopic areas of human primary visual cortex representing the visual stimulus (Watkins et al., 2006).

Such an association of V1 activity with illusory cross-modal perception is consistent with earlier findings that visual-evoked potentials and fields are modified at short latency in association with the illusion (Bhattacharya et al., 2002, Shams et al., 2001 and Shams et al., 2005). Moreover, it may suggest that activity in V1 reflects subjective perception rather than the visual stimulus that was physically presented. However, an alternate possibility is that enhanced V1 activity for the cross-modal 'fission' illusion might represent the effects of either fluctuating attention, or a non-specific response to a perceptual matching between sensory modalities, rather than a response that truly varied with perception.

Here, I sought to rule out this possibility by replicating my earlier findings and now comparing them with a complementary illusion (Andersen et al., 2004). In contrast to the previous chapter, which focused on situations where one physical flash was incorrectly perceived as two flashes ('fission'), here I focused on situations where two physical flashes are incorrectly perceived as one flash ('fusion'). If V1 activity reflects subjective perception, then it should be enhanced for the 'fission' but reduced for the 'fusion' illusion, reflecting the illusory perception of two (when one was physically present) or one (when two were physically present) flashes respectively. However, an account of the illusion that postulates V1 responses reflecting the modulatory influences of attention or auditory-visual matching predicts that V1 activity should be enhanced both for 'fission' and 'fusion' illusions.

## 6.2 Methods

### *6.2.1 Subjects*

Fourteen young adults (6 females, 18-30 years old, right handed) with normal hearing and normal or corrected to normal vision gave written informed consent to participate in the study, which was approved by the local ethics committee. Prior to scanning all subjects took part in a behavioural pilot experiment (see Procedure, below, for full details), following which two subjects were excluded because they did not report the multisensory illusion. Following scanning, two subjects were rejected on the basis of

excessive head movement (>5mm). Ten subjects (6 females, 18-30 years old, right handed) were therefore included in the analyses reported here.

### *6.2.2 Stimuli*

Visual stimuli were projected from an LCD projector (NEC LT158, refresh rate 60 Hz) onto a circular projection screen at the rear of the scanner. The subjects viewed the screen via a mirror positioned within the head coil. The auditory stimuli were presented binaurally using electrostatic headphones (KOSS, Milwaukee, USA. Model: ESP 950 Medical) custom adapted for use in the scanner. All stimuli were presented using MATLAB (Mathworks Inc.) and COGENT 2000 toolbox (www.vislab.ucl.ac.uk/cogent/index.html). Visual stimuli consisted of an annulus with luminance 420cd/$m^2$ and eccentricity 8.5 -10 degrees of visual angle presented for 17ms. When two flashes were presented, the interval between them was 46ms.The background was a uniform gray screen of luminance 30cd/$m^2$. Luminance calibration was achieved via a viewing aperture in the MRI control room using a Minolta LS-100 spot photometer. I used an annulus displayed in the peripheral visual field in association with auditory stimulation to maximise illusory perception, which is stronger for stimuli displayed in the periphery (Shams et al., 2002). In addition, the cortical representation of such a peripheral annulus avoids the foveal confluence at the occipital pole (Sereno et al., 1995), where it is extremely difficult to distinguish activity from different early retinotopic visual cortical areas. This stimulus geometry enabled me to clearly distinguish activity in V1, V2 and V3 from other cortical areas. The auditory stimuli consisted of a sine wave with frequency 3.5kHz, duration 10ms (with a ramp time of 1ms at each end of the sound wave envelope) and volume 95dB. The sound intensity (SPL) produced by the headphones was measured while the headphones were a suitable distance away from the scanner using a sound meter (Radioshack 33-2055). When two bleeps were presented, the interval between them (ISI) was 46ms. On trials with two flashes and one bleep, the auditory bleep was presented simultaneously with the first flash. Similarly, on trials with two bleeps and one flash the flash was presented simultaneously with the first bleep. Pilot behavioural work confirmed previous observations that whether bleeps and flashes are presented simultaneously or with slight temporal offset (Shams et al., 2002) makes little difference to behavioural reports of illusory perception.

### *6.2.3 Procedure*

Subjects initially took part in a behavioural pilot study. Subjects were presented with one or two briefly and successively flashed visual stimuli, either alone or accompanied by one or two successively presented auditory bleeps. For clarity, these trial types will be referred to by abbreviations. For example, 'F2B1' refers to trials on which there were two flashes and one bleep while 'F2B2' refers to a trial on which two flashes and two bleeps were presented. Subjects were instructed to report by button press whether they perceived one or two flashes and ignore the bleeps. Each participant completed 1 run of 128 trials divided between the different trial types (F1B1, F1B2, F2B1, F2B2, F1 & F2) while in the scanner. This pilot study ensured that the subjects experienced the multisensory illusion and could clearly distinguish one and two flashes with no auditory stimulation (all subjects were able to achieve >95 % correct on visual alone trials before starting the main experiment). The two subjects who were excluded showed no difference in error rates in the F1B1 and F2B2 conditions (5% & 4% error rates for each excluded subject, compared to 6 % (SD 1 %) for the group who went forward to the experiment.

During the fMRI experiment, on each trial subjects were presented with one or two briefly and successively flashed visual stimuli accompanied by one or two successively presented auditory bleeps. These comprised four different trial types that represented all the possible combinations of flashes and bleeps. Subjects maintained central fixation throughout and indicated whether they perceived one or two flashes, by pressing one of two response keys on a keypad held in their right hand. Each trial lasted 90ms followed by a 1800ms response interval. Eye position data was collected from all subjects during the trials to ensure subjects maintained fixation. One seventh of all trials were null trials, during which no visual or auditory stimuli were presented. There were thus five physically different types of trial. The responses of subjects were further used to post hoc divide the F2B1 and the F1B2 trials into those on which the illusion was perceived (" F2B1- Fusion Illusion" and "F1B2 - Fission Illusion"), and those on which it was not ("F2B1 - no Illusion" and "F1B2 - no Illusion"). Each subject completed between 4 and 6 runs of 112 trials divided equally between the different trial types. Trials were pseudo-randomly distributed within a run.

### *6.2.4 fMRI scanning*

A 3T Siemens Allegra system was used to acquire both T2*-weighted echoplanar (EPI) images with Blood Oxygenation Level Dependent contrast (BOLD) and T1 weighted anatomical images. Each EPI image comprised of thirty-two 3mm axial slices with an in-plane resolution of 3x3mm positioned to cover the whole brain. Data were acquired in four to six runs, each run consisting of 162 volumes. The first five volumes of each run were discarded to allow for T1 equilibration effects. Volumes were acquired continuously with a TR of 2.08s per volume. During scanning, eye position and pupil diameter were continually sampled at 60Hz using long-range infrared video-oculography (ASL 504LRO Eye Tracking System, Mass). Eye movements were monitored on-line via a video screen for all subjects. Subjects completed a short pilot in the scanner to ensure they could maintain fixation.

## 6.3 Data analysis

### *6.3.1 Eye tracking*

Eye tracking data were analysed with MATLAB (Mathworks Inc.). Blinks and periods of signal loss were removed from the eye movement data. Mean eye position, expressed as a distance from fixation, was then computed for each trial type and every subject from whom data were available. A repeated-measures ANOVA was used to establish whether mean eye position deviated significantly from fixation, or between conditions.

### *6.3.2 fMRI preprocessing*

#### *6.3.2.1 Spike artefacts*

The EPI magnitude images undergoing statistical analysis were reconstructed from the complex k-space raw data using a generalized reconstruction method based on the measured EPI k-space trajectory to minimize ghosting (Josephs O et al., 2000). Prior to reconstruction the k-space raw data were assessed for spike artefacts as indicated

by high background noise (two-fold oversampling in the readout direction always allowed for estimating the background noise from areas outside the head) (Weiskopf N et al., 2006). If k-space phase-encoding lines were affected by spikes they were replaced by the corresponding k-space lines from adjacent uncorrupted time points of the EPI time series. A correction for linear phase variations across k-space (due to inter-scan motion) was applied prior to replacing the data. Replacing single k-space lines instead of complete slices or volumes ensured that a minimal amount of data were interpolated. Less than 0.03% of all k-space lines required correction, thus minimally affecting the experimental degrees of freedom. The spike detection and correction were implemented in MATLAB (Mathworks Inc.).

#### *6.3.2.2 Preprocessing*

The resulting functional imaging data were analyzed using Statistical Parametric Mapping software (SPM2, Wellcome Department of Imaging Neuroscience, University College London). All image volumes were realigned spatially to the first, and temporally corrected for slice acquisition time (using the middle slice as a reference). Resulting image volumes were coregistered to each subject's structural scan. The fMRI data were analyzed using an event-related model. Activated voxels in each experimental condition for each subject were identified using a statistical model containing boxcar waveforms representing each of the experimental conditions, convolved with a canonical hemodynamic response function and mean corrected. Motion parameters defined by the realignment procedure were added to the model as six separate regressors of no interest. Multiple linear regression was then used to generate parameter estimates for each regressor at every voxel for every run. The resulting parameter estimates were averaged across runs to give a final parameter estimate for each of the experimental conditions for every subject. In order to get an accurate parameter estimate for each condition, any run with less than 4 events in a given illusion condition was excluded from the analysis. Data were scaled to the global mean of the time series and high pass filtered (cut-off: 0.0083 Hz) to remove low-frequency signal drifts.

### *6.3.3 Retinotopic analyses*

To identify the boundaries of primary visual cortex, standard retinotopic mapping procedures were employed (Sereno et al., 1995;Teo et al., 1997;Wandell et al., 2000). Only seven of the ten subjects participated in the retinotopic mapping procedures and so data from these seven are reported here, while the non-retinotopic analyses reported below used all ten subjects. There were no behavioural or demographic differences between the two groups. Flashing checkerboard patterns covering either the horizontal or vertical meridian were alternated with rest periods for 16 epochs of 26 s over a scanning run lasting 165 volumes. SPM2 was used to generate activation maps for the horizontal and vertical meridians. Mask volumes for each region of interest (left and right V1, V2, and V3) were obtained by delineating the borders between visual areas using activation patterns from the meridian localisers. Standard definitions of V1 were followed together with segmentation and cortical flattening in MrGray (Teo et al., 1997;Wandell et al., 2000). Using the mask volumes for left and right V1, V2 and V3, I identified voxels that showed significant activation ($p<.05$ uncorrected) for the comparison of all trials on which visual stimulation was present (i.e. all experimental conditions) compared to null events, employing the regression analysis described above. This comparison identifies voxels activated by the annular visual stimulus in each of the retinotopic areas. Informal examination of these activations superimposed on flattened representations of occipital cortex confirmed our expectation that they represented voxels activated by the annular visual stimulus (Figure 6.1).

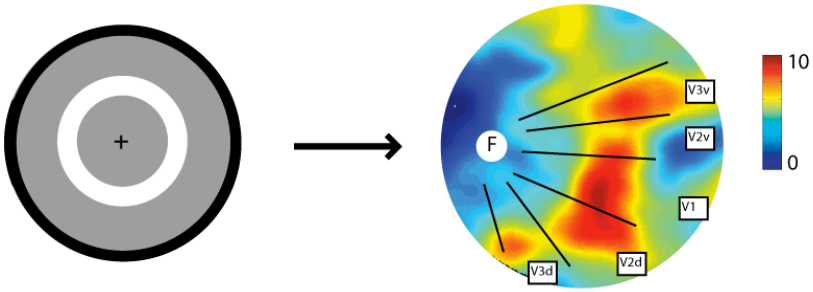


**Figure 6.1 Spatial distribution of stimulus-evoked activity in retinotopic visual cortex.** The statistical contrast of all visual events (**F1B1**, **F1B2**, **F2B1**, **F2B2)** versus null events thresholded at p<.05 uncorrected) is shown projected onto a flattened representation of visual cortex for a representative subject (the letter 'F' represents the location of the fovea, corresponding to the occipital pole; and the colourscale represents the t value at each location for the statistical contrast above, where red represents highest t values and blue the lowest; see Methods for full details).

Having thus independently identified the stimulus representation in V1-V3, I then extracted and averaged the regression parameters from the analysis of the main experimental time-series (described above). This procedure yielded estimates of percentage signal change for each condition averaged across voxels in V1, V2 and V3 that responded to the visual stimulus. The percentage signal change was divided by the average cortical response to visual stimulation (i.e. ([F1 +F2]/2) in each subject to produce a normalised percentage signal change for each condition. The statistical significance of any differences in activation between the Illusion condition and the No-Illusion conditions was assessed by entering the normalised percentage signal change for each subject in each condition into a two tailed t test using a significance level of p<.05. Finally, a mask representing the voxels in V1 that did not show a significant response to the visual stimulus was calculated. I then used this image to repeat the above procedure to examine the response to each condition in the non stimulus responsive area of V1.

### *6.3.4 Whole brain analysis*

To complement the retinotopic analyses, I also conducted an unbiased examination of regions outside retinotopic cortex using a random-effects whole-brain analysis of all ten subjects. The realigned and slice time corrected images from each subject were spatially normalized to a standard EPI template volume based on the MNI reference brain in the space of Talairach and Tournoux (Talairach Tournoux, 1988). The normalized image volumes were then smoothed with an isotropic 9mm FWHM Gaussian kernel. These data were analyzed using an event-related random-effects model, the first stage of which was identical to the regression model described above for the retinotopic analyses, except now applied to spatially normalised images. The parameter estimates for different conditions were then entered into a second level analysis using planned comparisons with paired t-tests. For these whole brain analyses, a statistical threshold of $p<.05$ corrected for multiple comparisons was used except for areas previously associated with the fission illusion where a small volume correction (sphere of diameter 3mm centered on coordinates [54 -54 30]) was applied (Watkins et al., 2006).

## 6.4 Results

### *6.4.1 Behavioural*

Analysis of behavioural responses during scanning confirmed that subjects were able to accurately report the number of flashes when the number of flashes and bleeps were identical (i.e. F1B1 and F2B2 trial types; accuracy 94%, SE across subjects 1%). On a large proportion of trials when two flashes were accompanied by one bleep (F2B1 trials), subjects reported an illusory perception of one flash ("F2B1- Fusion Illusion"; 42% of all F2B1 trials, SE across subjects 6%). On the remainder of F2B1 trials, subjects reported veridical perception of one flash ("F2B1-no Illusion"). When one flash was accompanied by two bleeps (F1B2) subjects reported an illusory perception of two flashes (F1B2 – Fission illusion) on 34% of the trials (SE across subjects 7%). The frequency of occurrence of the fusion illusion was not significantly different from the Fission illusion (42 % versus 44% $t(9) = 1.24$, $p = 0.25$).

### *6.4.2 Eye position data*

Subjects were requested to maintain fixation at the centre of the display. During scanning eye position was monitored on-line in all subjects to ensure subjects successfully maintained fixation throughout the experiment sessions. A repeated-measures ANOVA showed no statistical difference in mean eye position from fixation, or between conditions for the eight subjects in whom eye data were available ($F(6,42) = .957$, $p = .466$). Eye data was monitored but not recorded in two subjects.

### *6.4.3 Functional MRI*

#### *6.4.3.1 Retinotopic analyses*

Many F2B1 trials (42%) evoked the illusion of one flash (F2B1- Fusion Illusion), while on the remainder two flashes were perceived (F2B1-no Illusion). I therefore compared activity in retinotopic visual areas that was evoked on F2B1- Fusion Illusion trials with F2B1-no Illusion trials and on F1B2- Fission Illusion trials with F1B2-no Illusion trials in the seven subjects where retinotopic maps were obtained. Stimulus-evoked activity in V1 was significantly lower for F2B1-Fusion Illusion trials on which the illusion was perceived, compared to F2B1-no Illusion when the illusion was not perceived ( [$t(6) = 2.93$, $p = .026$], two-tailed) (see Figure 6.2 for full details, including time courses). Note that I compared physically identical F2B1 trials with exactly the same visual and auditory stimulation that resulted either in the fission illusion or no illusion. Thus, any differences in brain activity associated with this comparison *cannot* reflect differences in visual or auditory stimulation. Stimulus-evoked activity in V1 in the F2B1- Fusion Illusion condition (where one flash was perceived) was not significantly different from the F1B1 condition (where one flash was physically present) [$t(6) = .79$; $p = .45$] (see Figure 6.2 for full details, including time courses).

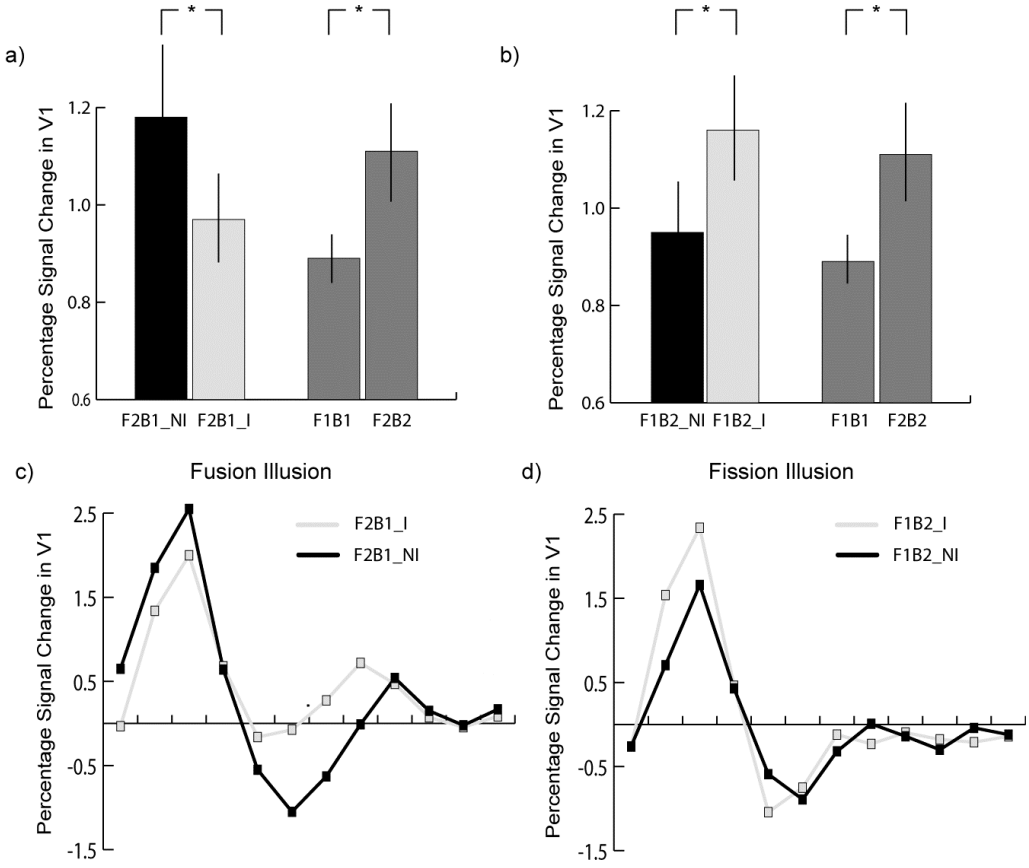


**Figure 6.2 Signal change in primary visual cortex associated with illusory multisensory perception.**

a) The mean percentage signal change in retinotopically defined V1 (see Methods) is shown for the condition **F2B1-no Illusion** (two flashes with one bleep when subjects reported correctly the perception of two flashes), **F2B1-Fusion Illusion** (two flashes with one bleep when subjects reported the illusory perception of one flash), **F1B1** and **F2B2**. b) The mean percentage signal change in retinotopically defined V1 is shown for the condition **F1B2-no Illusion** (one flash with two bleeps when subjects reported correctly the perception of one flash), **F1B2- Fission Illusion** (one flash with two bleeps when subjects reported the illusory perception of two flashes), **F1B1** and **F2B2**. Data shown are averaged across the seven subjects (see Methods for further details) with error bars representing the standard error of the mean, and the symbol '*' indicating statistical significance ($p<.05$). c) Time courses for the V1 cortical responses in the **F2B1- Fusion Illusion** (grey line) and **F2B1-no Illusion** (black line) condition. d) Time courses for the V1 cortical responses in the **F2B1-Fission Illusion** (grey line) and **F2B1-no Illusion** (black line) condition. Percentage signal change in V1 is plotted against time from stimulus onset (units of TR = 2.08 seconds) for both conditions averaged across subjects. The time courses were calculated for each of the subjects by using a statistical model containing a boxcar waveform representing each of the experimental

conditions, convolved with a series of FIR (finite impulse response) functions (using the SPM toolbox MarsBaR; http://marsbar.sourceforge.net/). Motion parameters defined by the realignment procedure were added to the model as six separate regressors of no interest. Multiple linear regression was then used to generate parameter estimates for each regressor at each time point for every subject. The data used in this model were extracted from the area of V1 that responded to the visual stimulus. This was determined by masking the entire V1 region of interest (see Methods) with the cortical area that showed significant responses ($p<.05$ uncorrected) to the contrast of all visual events (F1B1, F1B2, F2B1 & F2B2) versus null events.

Similarly, many F1B2 trials (34%) led to the illusion of two flashes (F1B2- Fission Illusion), while on the remainder only one flash was perceived (F1B2-no Illusion). The stimulus-evoked activity in V1 was significantly higher for F1B2-Fission Illusion trials on which the illusion was perceived compared to F1B2-no Illusion when the illusion was not perceived ([$t(6) = 2.70$, $p = .035$], two-tailed) (see Figure 6.2). The activity in V1 in the F1B2-Illusion condition (where two flashes where perceived) was not significantly different from the F2B2 condition (where two flashes where physically present) [$t(6) = .32$; $p = .75$]. This replicates my previous findings (Watkins et al 2006). Again, because physically identical trials are compared these differences cannot be attributed to differences in sensory stimulation. The activity in stimulus-driven regions of V2 and V3 showed a similar pattern of activation to V1 but did not reach conventional levels of statistically significance (Figure 6.3)

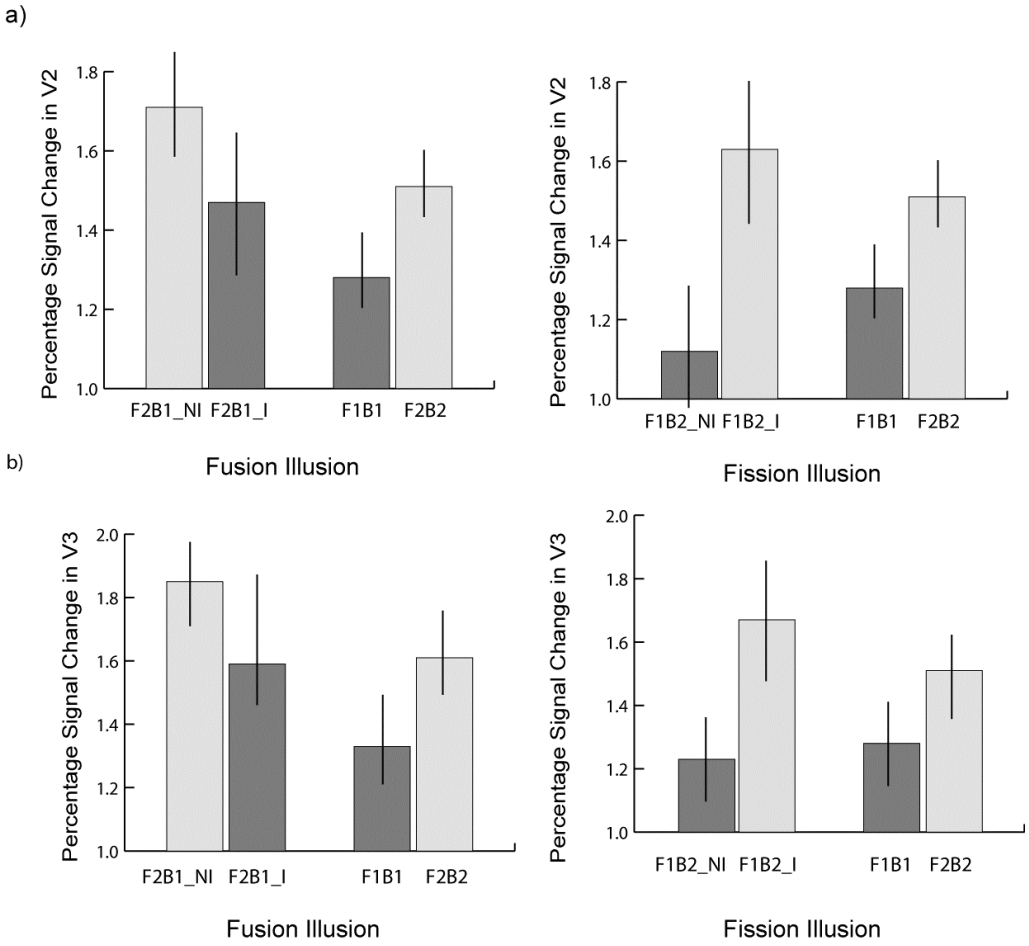


**Figure 6.3 Signal change in V2 & V3 associated with illusory multisensory perception.** a) The mean percentage signal change in retinotopically defined V2 (see Methods) is shown for the condition **F2B1-no Illusion** (two flashes with one bleep when subjects reported correctly the perception of two flashes), **F2B1- Fusion Illusion** (two flashes with one bleep when subjects reported the illusory perception of one flash), **F1B2-no Illusion** (one flash with two bleeps when subjects reported correctly the perception of one flash), **F1B2- Fission Illusion** (one flash with two bleeps when subjects reported the illusory perception of two flashes), **F1B1** and **F2B2**. b) The mean percentage signal change in retinotopically defined V3 (see Methods) is shown for the condition **F2B1-no Illusion**, **F2B1- Fusion Illusion**, **F1B2-no Illusion**, **F1B2- Fission Illusion**, **F1B1** and **F2B2**. Data shown are averaged across the seven subjects (see Methods for further details) with error bars representing the standard error of the mean.

These differential cortical responses to the fission and fusion illusory perception were specific to the retinotopic locations of V1 responding to the visual annulus, as there was no significant effect of the illusions in the regions of V1 that did not respond to

the visual annulus [Fission: t(6) = .80,p =.46 Fusion: t(6) = .83,p =.45 ]. Similarly, there was no evidence for a general effect of either judgment or number of flashes outside regions of V1 responsive to the visual annulus. Specifically, there was no significant difference between activity evoked in locations of V1 that did not correspond to the visual annulus for the judgment of 1 flash (i.e. conditions F1B1, F1B2 - no Illusion and F2B1 - Fusion Illusion) versus 2 flashes (i.e. conditions F2B2, F2B1 - no Illusion & F1B2 – Fission Illusion) or the actual presence of 1 flash (F1B1& F1B2) versus 2 flashes (F2B1 & F2B2) [t(6) = .21,p =.91, t(6) = 2.1,p = .08] Finally, there were no significant differences between the F1B1 and F2B2 conditions in the non-stimulus responsive area of V1 [t(6) = 1.48, p =.19].

#### *6.4.3.2 Whole brain analyses*

To complement the retinotopic analyses, whole-brain analyses of activity for each of the main comparisons outlined above was also performed. Unrestricted whole-brain analysis of illusory multisensory fusion perception (i.e. F2B1-Fusion Illusion vs F1B2-no Illusion) revealed significant activation in the right superior temporal sulcus ([58 -32 20]; t=7.30; p = .01 corrected at cluster level, number of voxels in the cluster = 96). These activated loci are shown in Figure 6.3. There were no cortical areas that showed a significant response to F1B2-no Fusion Illusion >F1B2 Fusion Illusion. Unrestricted whole-brain analysis of illusory multisensory fission illusion (i.e. F1B2 – Fission Illusion vs F1B2-no Illusion) revealed no significant cortical activation outside early visual areas at a corrected threshold. An examination of cortical areas previously associated with this illusion revealed significant activation in the right superior temporal sulcus ([52 -54 28]; t =2.9; p =0.04, corrected for small volume examined).

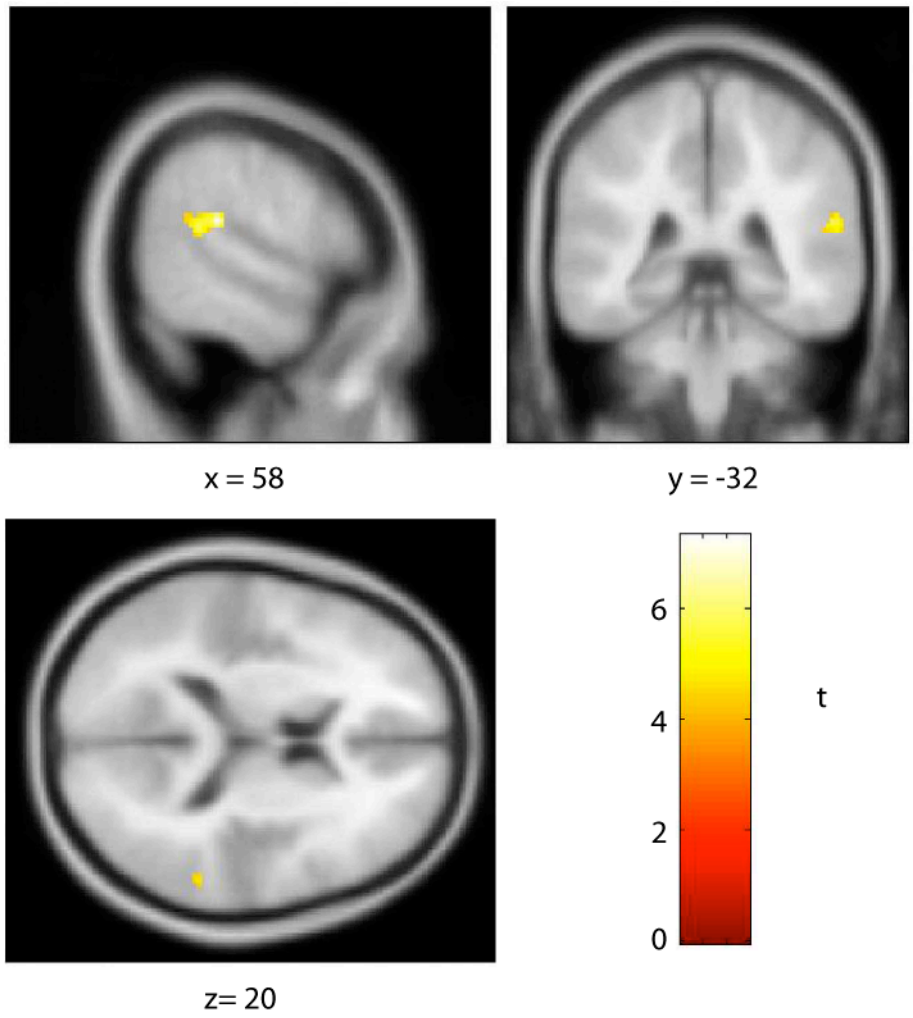


**Figure 6.4 Cortical areas activated by multisensory illusory perception outside retinotopic cortex.**

Shown in the figure are cortical loci outside retinotopic cortex where event-related activity was significantly greater during **F1B2-Fusion Illusion** trials compared to **F1B2-no Illusion** trials ($p<.05$, corrected for multiple comparisons; see also Results). Activated cortical loci in the right superior temporal sulcus projected onto a T1 template image in the stereotactic space of Talairach & Tournoux (1988).

## 6.5 Discussion

These behavioural findings demonstrated that subjects perceived an illusory perception of one flash ('fusion') rather than the veridical perception of two flashes on many F2B1 trials. We found that brain activity evoked in human V1 on these fusion illusion trials (F2B1-Fusion Illusion) was significantly lower than on physically identical trials where no illusion was reported (F1B2-no Illusion). In agreement with previous findings (Bhattacharya et al., 2002;Watkins et al., 2006) I also demonstrated that activity on fission illusion trials (F1B2-Fission Illusion) was significantly higher than on physically identical trials where no illusion was seen (F2B1- no Illusion). This modulation of activity in association with illusory perception did not reflect differences in eye position or eye movements on different trials. Thus, perception of either the fission or fusion illusion caused opposite effects on activity in primary visual cortex. When two flashes were presented but one perceived, activity was decreased; but when one presented and two perceived, activity was increased. The level of cortical activity in V1 was therefore associated with conscious visual perception rather than the physically present stimulus. The modulation of activity by illusory multisensory perception was only found in the stimulus responsive area of primary visual cortex. This demonstrates that an auditory effect in primary visual cortex is specific to the area representing the visual stimulus, and reflects a modulatory influence on visual stimulation.

Critically, these divergent effects on V1 activity that follow perception cannot be explained by a general attentional effect, nor a response of early visual areas to a match between physically present stimuli and perception. For the fission multisensory illusion reported here, visual evoked potentials and fields associated with the illusory perception are modified at a short latency (Bhattacharya et al., 2002;Shams et al., 2001;Shams et al., 2005) consistent with generators in early visual cortex (although note that temporally early effects reported from earlier studies do not *necessarily* translate into anatomically early effects such as generators in the early visual cortex). In addition the previous chapter has demonstrated increased cortical activity in V1 in association with illusory perception of an additional visual flash (Watkins et al, 2006). These new findings extend this earlier work by demonstrating conclusively that activity in V1 follows multisensory perception.

The general finding that V1 activity can be more closely related to conscious visual experience rather than physical stimulation is recognised in unisensory studies. For example, activity evoked in human V1 by a visual stimulus briefly presented at the contrast detection threshold is higher on trials when subjects successfully detect it than when they fail to do so. Moreover, when subjects falsely perceive the presence of a low-contrast stimulus on trials when the stimulus was physically absent (false alarms), V1 activity is similar to that on trials where subjects correctly report the physical presence of a stimulus (Ress and Heeger, 2003). The present findings show that such an association of V1 activity with conscious perception extends to suprathreshold visual stimuli and to changes in visual perception brought about by multisensory stimulation.

These data do not precisely define how the association of V1 activation with illusory visual perception occurs, nor whether the modulation of V1 activity we observed plays a causal role in the generation of the illusion. Primary visual cortex receives projections from at least twelve areas belonging to the visual cortex (Felleman and Van Essen, 1991). Recently more distant projections have been described from areas in the ventral (Distler et al., 1993b) and dorsal visual pathways and from the lateral intraparietal area (Boussaoud et al., 1990;Rockland and Van Hoesen, 1994). Several recent papers have used tracer injections to demonstrate projections from primary auditory cortex, auditory association areas and the superior temporal polysensory area (STP) to the area of primary visual cortex representing the peripheral visual field (Clavagnier et al., 2004;Falchier et al., 2002;Rockland and Van Hoesen, 1994). The function of these projections to V1 has been the subject of much debate they may serve to enhance perceptual capabilities; for example the addition of an auditory signal to a visual signal leads to improved detection compared to a visual signal alone (Bolognini et al., 2005;Frassinetti et al., 2002;Gondan et al., 2005;Miller, 1982;Molholm et al., 2002;Schroger and Widmann, 1998b). Thus, it is possible that these direct connections mediate the changes in V1 activity that I observed. This is consistent with previous findings showing that multisensory influences also extend to the earliest stages of cortical processing (Ghazanfar and Schroeder, 2006). Interestingly, recent studies examining somatosensory-auditory multisensory integration in primary auditory cortex show that auditory input to A1 occurs via

feedforward projections from the thalamus and it is possible that similar low level thalamic connections may extend to auditory-visual multisensory integration (Lakatos et al., 2007).

When subjects experienced the fusion illusion, activity was increased in the right superior temporal sulcus (STS) and decreased in primary visual cortex. I have thus now found evidence that the right STS is involved in both the fission and fusion illusions. The area of the right STS activated in the fission illusion is posterior to the cortical area involved in the fusion illusion. However, the size of the clusters and spatial smoothness of our data mean that it is not at present clear whether these activations reflect two distinct cortical loci. However, as regions of the right STS show a similar response for two illusions that are both perceptually very different and exhibit very different activation patterns in V1, these data suggests that the right STS may not be playing a causal role in generating the illusory perception. The low temporal resolution of fMRI signals mean that we cannot determine whether the STS activation we observed was casually related to the changes in V1 activity, or a later effect. However, we speculate that the STS response may occur later and represent a response to the matching of auditory and visual perception (i.e. the perception of F1B2 Fission illusion would be effectively F2B2 compared to the non illusion perception of F1B2). Such a speculation would be consistent with both the divergent effects of the fusion and fission illusions on activity in primary visual cortex, and previous studies demonstrating early audiovisual integration(Bhattacharya et al., 2002;Giard and Peronnet, 1999;Shams et al., 2001). The STS has been consistently associated with integration between visual and auditory stimulation (Barraclough et al., 2005;Beauchamp et al., 2004;Beauchamp, 2005;Calvert et al., 2000;Ghazanfar et al., 2005;Olson et al., 2002;Schroeder and Foxe, 2002) Interestingly, a recent study has shown that the STS is involved in multisensory associative learning (Tanabe et al, 2005). Further research will be needed to elucidate the precise role of the STS in these multisensory illusions.

### 6.6 Conclusion

In this chapter I have demonstrated that fMRI signals from stimulus-responsive regions of human primary visual cortex closely corresponded to multisensory

perception for both ‘fission’ and ‘fusion’ illusions. Moreover, when auditory stimulation gave rise to an illusory change in perceptual experience this was associated with increased activity in the right superior temporal sulcus.

# CHAPTER 7: SOUND MOVES LIGHT: PSYCHOPHYSICAL AND FMRI EVIDENCE OF AUDITORY-DRIVEN VISUAL APPARENT MOTION.

## 7.1 Introduction

In the previous two chapters I have demonstrated that the activity in primary visual cortex closely follows multisensory perception in the flash-bleep illusion. In this chapter I examine a different type of multisensory illusion called 'temporal ventriloquism'.

In the 'temporal ventriloquism' effect, an auditory event that either leads or lags a visual event can appear to shift the visual onset backwards or forwards in time respectively (Gebhard et al., 1959;Morein-Zamir et al., 2003;Vroomen et al., 2004). Recent work (Freeman and Driver, 2008) has demonstrated that pure auditory timing can influence visual processing of spatio-temporal patterns, namely motion. A long-debated issue concerns whether such phenomena reflect feedforward convergence of multimodal timing information in higher cortical areas, or alternatively crossmodal interactions between sensory processing in early sensory cortices (Ghazanfar and Schroeder, 2006). In this chapter I used functional MRI with multivariate pattern classification to demonstrate that the direction of visual apparent motion can be predicted from patterns of activation in motion-responsive visual areas (V3 and MT+). Remarkably, such patterns in V3 and MT+ can also be used to predict the perceived direction of ambiguous visual apparent motion whose perceived direction is biased solely on auditory information.

In classical visual apparent motion bars flashed alternately on the left and right of a visual display appear to sweep rapidly back and forth across the screen (von Grunau, 1986). A recently described crossmodal illusion has demonstrated that the timing of an auditory stimulus that accompanies the visual stimulus can bias the perceived direction of visual apparent motion (Freeman and Driver, 2008). This behavioural effect provides further support to the hypothesis that visual processing is subject to modulatory influences from other sensory modalities (Ghazanfar and Schroeder, 2006) similar to other reported cases of auditory influences on visual processing (Andersen et al., 2004;Berger et al., 2003;Bhattacharya et al., 2002;Ghazanfar et al.,

2005;McGurk and MacDonald, 1976;Muckli et al., 2005;Shams et al., 2000;Shams et al., 2002;Watkins S et al., 2007;Watkins et al., 2006). To test this hypothesis further, I examined cortical responses to auditory-driven visual apparent motion, using functional magnetic resonance imaging (fMRI).

In this study, I manipulated the asynchrony of a bleep accompanying visual bars which alternated regularly between left and right positions on the screen (see Fig 7.1a and 7.1b). Rightwards motion typically dominates when one bleep lags the left flash and another leads the right flash, so that the interval between bleeps paired to right and left flashes was shorter than for the left to right flashes (with leftwards motion dominating for the bleeps leading the left and lagging the right flashes). Note that the regularly flashing visual stimulus provides no directional motion cues, and thus the direction of perceived motion is driven purely by the timing of the auditory stimulus (Freeman and Driver, 2008). I compared this with the more typical form of visually-driven apparent motion where the perception of motion is induced in silence by the stimulus onset asynchrony between the left and the right visual stimulus (von Grunau, 1986) (e.g. rightwards motion perceived when the interval between left-to-right flashes is shorter than for right-to-left flashes, Fig 1). In the previous study auditory and visual timing were both equally effective at determining the direction of apparent motion, and both also induced a robust visual motion after-effect.

In fMRI studies of visual apparent motion, activation has been observed in MT+ when motion is perceived compared to flicker and also in primary visual cortex along the path of apparent motion (Muckli et al., 2005;Sterzer et al., 2006). However the methods used in those studies do not provide a measure of the selectivity or specificity of activations for particular motion directions. Such directional selectivity has been measured in fMRI, but for continuous or 'short-range' visual motion, where 'repetition suppression' was observed when the same direction of motion was repeatedly shown on successive trials (Huk and Heeger, 2002;Seiffert et al., 2003). However, recent pattern classification methods offer a more direct measure of motion direction selectivity For attended smoothly-moving random dot kinematograms (Kamitani and Tong, 2006), activity patterns in human visual cortex (both early retinotopic areas V1-V3 and motion area MT+) have been found to contain robust

direction-selective information, from which it is possible to decode the direction of motion that is seen.

No studies to date have identified the distinct neural correlates of perceiving leftwards versus rightwards directions of long-range apparent motion. In this study I use multivariate pattern classification (see methods for further information) to test whether the perceived direction of visual apparent motion can be reliably decoded from early visual areas.

In the case of auditory-driven apparent motion, traditionally 'unimodal' visual areas might not be expected to respond differently to different directions of auditory-driven apparent motion. After all, the local visual stimulus contains no visual spatio-temporal information that could bias perception consistently towards leftwards or rightwards motion.

## 7.2 Methods

### *7.2.1 Subjects*

Eight healthy young adults (4 females, 18-35 years old, one left-handed) with normal hearing and normal or corrected-to-normal vision gave written informed consent to participate in the study. Procedures were approved by the local ethics committee.

### *7.2.2 Stimuli*

Visual stimuli were back-projected from an LCD projector (NEC LT158, refresh rate 60 Hz) onto a circular projection screen at the rear of the magnet bore. The subjects lay supine in the scanner and viewed the screen via a mirror positioned within the head coil from a distance of 62cm. The auditory stimuli were presented binaurally using electrostatic headphones (ESP 950 Medical, KOSS, Milwaukee, USA) custom adapted for use in the scanner. On-line eye-tracking (60 Hz) was provided by a long-range infra-red video camera (ASL 504LRO, Applied Science Laboratories, Massachusetts, USA ) trained on the left eye. All stimuli were presented using MATLAB (Mathworks Inc.) and COGENT 2000 toolbox

(www.vislab.ucl.ac.uk/cogent/index.html). Subjects made responses via key-presses on a custom-built MRI compatible button-box.

Visual stimuli comprised peripherally viewed vertical red (CIE chromaticity coordinates x=.629, y=.348) bars with luminance of 30 $cdm^{-2}$ presented in alternation on the left or right of the vertical midline on a black background. A white fixation point (diameter 0.4 deg, luminance 120 $cdm^{-2}$) was displayed 13 degrees below the centre of the display before and during each trial (see Figure 7.1a). The auditory signal was a 60ms rectangular-windowed 480Hz sine-wave, sampled at 22kHz with 8-bit quantization. The volume of this auditory stimulus was adjusted for each subject so that signals were clearly audible above the scanner noise.

At the start of each trial, both bars were presented simultaneously in silence for 500ms. All subsequent visual and auditory events were programmed relative to a regular 1Hz event-cycle, which commenced with the offset of one randomly selected bar. Commencing the alternating sequence with a random offset helped to balance any bias towards perceiving motion in a direction initially away from the first onset. Each 1000ms event-cycle consisted of an L and R bar each appearing in alternation for 200ms (Fig. 7.1a). In the Visual Timing sequence, the SOA between L and R bars ($vSOA_{LR}$) was set to either 500-166ms for the appearance of unidirectional rightwards apparent motion, or to 500+166ms for leftwards apparent motion (see Figs. 7.1b-c; note that SOA between R and L was always reciprocal, such that $vSOA_{RL}$ + $vSOA_{LR}$ = 1000ms). For the Auditory Timing sequence, vSOA was set to 500ms (Fig. 7.1d) and auditory signals of 60ms were each paired asynchronously with left and right flashes. When $aSOA_{LR}$ (the interval between sounds paired to left and right flashes) was equal to 500ms-166ms (Fig. 7.1e), one bleep lagged the onset of the left bar by 83ms, while another bleep preceded onset of the right bar by the same interval. This could induce rightwards visual apparent motion. The opposite relationship was the case for $aSOA_{LR}$ = 500-166ms (Fig. 7.1f), inducing leftwards motion. Therefore there were four conditions: Visual Timing Right (Fig. 7.1b), Visual Timing Left (Fig. 7.1c), Auditory Timing Right (Fig. 7.1d) and Auditory Timing Left (Fig. 7.1e).

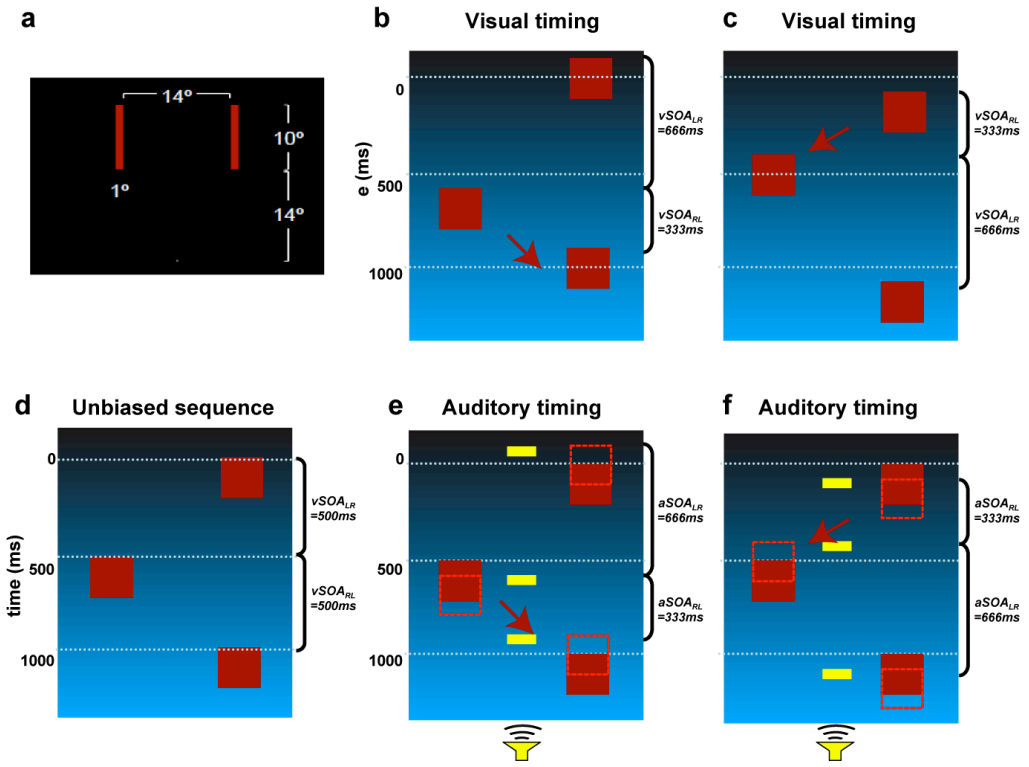


**Figure 7.1 Visual and auditory stimuli**

**(a)** Visual stimuli were peripherally viewed vertical red bars presented in alternation on the left (L) or right (R) side of the vertical midline on a black background. **(b-f)** Schematic representation of the Visual Timing condition. Space is represented horizontally and time vertically, with red bars representing visual events, and yellow bars auditory events. **(b)** Visual Timing sequence with Stimulus Onset Asynchrony (SOA) between R and L bars longer than for the return L-to-R direction (i.e. $vSOA_{RL} > vSOA_{LR}$), typically inducing unidirectional rightwards apparent motion (see red arrow); **(c)** the complementary case of $vSOA_{RL} < vSOA_{LR}$ for leftwards apparent motion. **(d)** Unbiased visual sequence with equal intervals between L-to-R and R-to-L flashes. **(e)** Auditory Timing sequence, with same equal visual intervals as in (d) but now with two auditory signals (see white bars), one lagging the onset of the left bar by 83ms, the other sound preceding onset of the right bar by the same interval. The shift in apparent visual onset via temporal ventriloquism is illustrated schematically by transparent orange bars. Red arrows indicate the rightwards direction of apparent motion resulting from these illusory temporal shifts. **(f)** The opposite timing relationship designed to induce leftwards apparent motion. Adapted from Freeman and Driver, 2008.

### *7.2.3 Procedure*

As part of the familiarization procedure outside the scanner, all subjects were shown silent sequences with an initially gross SOA asymmetry, and told that these displays could sometimes appear to have either leftwards or rightwards motion. Examples were then given of the Audio Timing condition. Without further prompting, all subjects could then readily discriminate between SOAs consistent with either L-R or R-L apparent motion. Training quickly progressed to smaller SOA differences, with minimal post-trial feedback. Further familiarization was conducted for approximately five minutes within the scanner, during which a range of five visual and auditory SOAs was presented in randomly-ordered blocks of 30 seconds.

In the main experiment, scanning sessions commenced with a blank fixation display for 13 seconds. There then followed eight consecutive blocks of 30 seconds duration, with an intervening rest period of 7 seconds (see Fig. 7.1 for description of these conditions). Visual or Auditory timing conditions were presented in pseudo-random order, with the initial direction counterbalanced over scanning runs and subjects. In each run there were 2 presentations of the 4 conditions (Visual Timing Left, Visual Timing Right, Auditory Timing Left and Auditory Timing Right). There were eight scanning runs for the first seven subjects and seven runs for the final subject, with time allowed for subjects to rest between runs.

### *7.2.4 Instructions to subjects*

Subjects were instructed to maintain constant fixation throughout the duration of a block, and to indicate the direction of motion that they perceived by pressing one of two keys on the button-box as soon as they had decided, using their preferred hand. If their percept changed then subjects were to switch response keys accordingly. No feedback was given during the main experiment.

### 7.2.5 fMRI data acquisition

A 3T Siemens Allegra system was used to acquire both T1 weighted anatomical images and T2*-weighted echoplanar (EPI) images with Blood Oxygenation Level

Dependent contrast (BOLD). Each EPI image volume comprised of forty 3mm axial slices with an in-plane resolution of 3x3mm positioned to cover the whole brain. Data was acquired in four runs, each consisting of 205 volumes. The first five volumes of each run were discarded to allow for T1 equilibration effects. Volumes were acquired continuously with a TR of 2.6s per volume.

## 7.3 Data analysis

### *7.3.1 fMRI preprocessing*

Functional imaging data were analyzed using Statistical Parametric Mapping software (SPM2, Wellcome Department of Imaging Neuroscience, University College London). All image volumes were realigned spatially to the first. Resulting image volumes were coregistered to each subject's structural scan. Activated voxels in each experimental condition for each subject were identified using a statistical model containing boxcar waveforms representing each of the experimental conditions, convolved with a canonical hemodynamic response function and mean corrected. Multiple linear regression was then used to generate parameter estimates for each regressor at every voxel for every subject. Data were scaled to the global mean of the time series and high pass filtered (cut-off: 0.0083 Hz) to remove low-frequency signal drifts.

### *7.3.2 Retinotopic analyses*

To identify the boundaries of primary visual cortex, standard retinotopic mapping procedures were used (Sereno et al., 1995;Teo et al., 1997;Wandell et al., 2000). Flashing checkerboard patterns covering either the horizontal or vertical meridian were alternated with rest periods for 16 epochs of 26 s over a scanning run lasting 165 volumes. SPM2 was used to generate activation maps for the horizontal and vertical meridians. Mask volumes for each region of interest (left and right V1, V2d, V2v, V3d, V3v) were obtained by delineating the borders between visual areas using activation patterns from the meridian localisers. We followed standard definitions of V1 together with segmentation and cortical flattening in MrGray (Morein-Zamir et al., 2003;Teo et al., 1997;Tootell et al., 1995;Vroomen and de Gelder, 2004;Wandell

et al., 2000). A separate motion localizer was used to functionally define MT+. A concentrically expanding and contracting radial dot field was alternated with a static dot field for 10 epochs of 22s over a scanning run lasting 105 volumes

### *7.3.3 Functional masks*

To identify visually responsive voxels to be used in the pattern classification, in appropriate retinotopic regions for the stimuli, functional masks were obtained. Using the mask volumes for left and right V1, V2, V3 and MT+ I identified 100 voxels (50 for MT+) that showed the most reliable activation in a t test comparing all trials on which a localizer stimulus was present compared to absent. This comparison identifies voxels activated by the stimuli in each of the retinotopic areas, independently of any differences due to perceived direction of motion.

### *7.3.4 Overview of multivariate analysis*

Prior to describing the details of how multivariate analysis was implemented in this study, I will briefly review how multivariate analyses work in the context of fMRI data. Consider the example shown in Figure 7.2. Two visual stimuli (images 1 and 2) evoke overlapping response patterns in visual cortex. These response patterns are sampled at low spatial resolution using fMRI to give a set of voxels that exhibit two sets of responses (one for each visual stimulus – figure 7.2a). The basis of multivariate analysis methods involves testing whether the response patterns evoked by the two stimuli are the same or different. If the response patterns to the two stimuli are different then the brain area under study can be said to encode the stimulus feature that varies between the two stimuli, whereas if the converse is true, this cannot be said to be the case. In univariate analyses each voxel is considered separately, therefore if differential neuronal responses to image 1 and 2 are distributed over a wide area, the difference in the response to each image at each voxel will be small. The advantage of multivariate analyses is that information is accumulated from all voxels enabling any differences in the pattern of activity evoked by image 1 and 2 to be compared. Multivariate analysis therefore allows assessment of whether the pattern of activity evoked by each stimulus can be accurately differentiated.

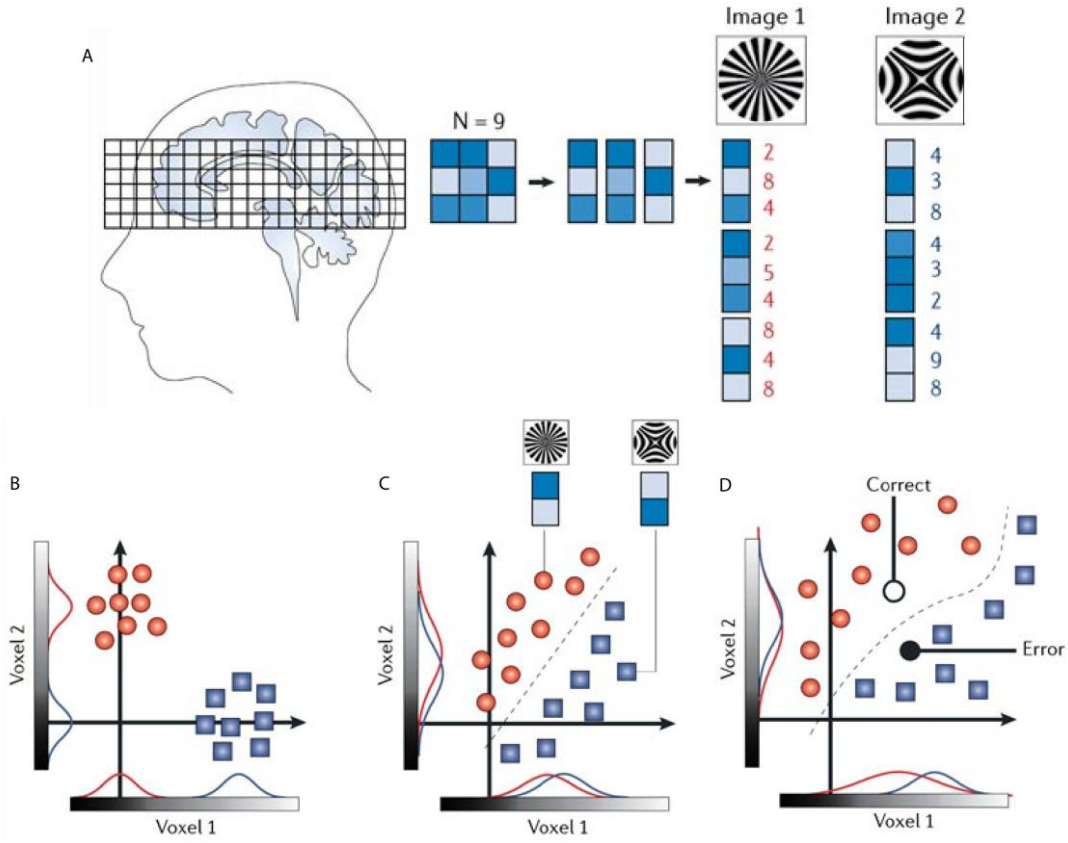


**Figure 7.2 Analysis of spatial patterns using a multivariate pattern recognition approach.**

**A**. FMRI measures brain activity repeatedly every few seconds in a large number of voxels (left). The joint activity in a subset (N) of these voxels constitutes a spatial pattern that can be expressed as a pattern vector (right). Different pattern vectors reflect different mental states; for example, those associated with different images viewed by the subject.

**B-D**. Each pattern vector can be interpreted as a point in an N-dimensional space (shown here in panels b–d for only the first two dimensions, red and blue indicating the two conditions). Each measurement of brain activity corresponds to a single point. A successful classifier will learn to distinguish between pattern vectors measured under different mental states. In panel b, the classifier can operate on single voxels because the response distributions (red and blue Gaussians) are separable within individual voxels. In panel c, the two categories cannot be separated in individual voxels because the distributions are largely overlapping. However, the response distributions can be separated by taking into account the combination of responses in both voxels. A linear decision boundary can be used to separate these two-dimensional response distributions. In panel d, to test the predictive power of a classifier, data are separated into training and test data sets. Training data (red and blue symbols) are used to train a classifier, which is then applied to a new and independent test data set. The proportion

of these independent data that are classified either correctly (open circle, 'correct') or incorrectly (filled circle, 'error') gives a measure of classification performance. This figure is adapted from (Haynes and Rees, 2006).

Practically this is achieved by training a pattern classification algorithm with the pattern of activity evoked by each image type. The classifier then attempts to blindly classify subsequently acquired test measurements to the category that evoked the most similar response pattern during the training phase (figure 6.5c). Accuracy is assessed by the comparing the classifier output with known labels (figure 6.5d). The steps required in this multivariate analysis are schematically represented in Figure 7.3.

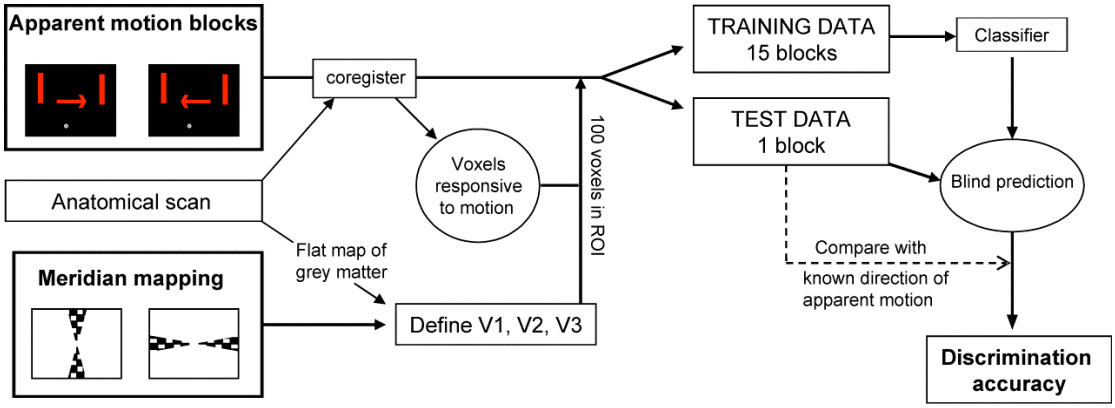


**Figure 7.3 Schematic representation of the steps in multivariate analysis**
All voxels activated by the audiovisual stimulus (irrespective of direction of motion) are identified (see methods for details). **B)** The raw fMRI signal over the whole experiment is extracted from each voxel in the stimulus representation to create a pattern vector of n voxels and their timecourses. The pattern vectors are split into two subsets, one for responses during leftwards motion and the other for rightwards motion. These vectors are then used as training sets for a pattern classification algorithm. The classifier then attempts to classify test measurements to the category that evoked the most similar response pattern during the training phase. The accuracy of this allocation of test measurements gives a measure of whether the pattern of activity evoked by each experimental condition can be accurately determined.

### *7.3.5 Pattern classification*

Based on the protocols for multivariate analysis outlined previously (Haynes and Rees, 2005), I extracted, for each participant, the raw BOLD signal from the voxels in each of V1, V2, V3 and V5/MT+ that were activated by the visual or audiovisual stimulus. The voxels were selected using the functional masks described above, so that they came from appropriate retinotopic regions responsive to the stimulus.

A total of 64 blocks of data were acquired for each subject, divided equally between the Auditory Timing (16 leftwards, 16 rightwards) and Visual Timing (16 leftwards, 16 rightwards) conditions. Visual and audiovisual apparent motion were analysed separately.

Each block contained 11 scans. The first 5 volumes from each block were discarded to allow the delayed BOLD signal to approach stability. The remaining 6 scans from each of the 15 blocks for each direction of apparent motion were assigned to a *training* data set, while the remaining images (the images from the remaining block) were assigned to be the *test* data set. Classification performance was then assessed using linear discriminant analysis with m-fold cross validation (Duda et al., 2001). Note that training and test datasets were from independent blocks, and that I used the raw unsmoothed fMRI signal in the realigned images (the statistical model was used only for identifying the voxels to be included in the analysis, as described above). This cycle was repeated 16 times using different blocks for training and test, and mean accuracy calculated over all 16 cycles. On each cycle, all voxels representing the stimulus were first rank ordered according to their T-value for the difference between the two directions of apparent motion. Then the *n* voxels with the highest T values were entered into the classification algorithm; this was repeated for a range of *n* (see below), with discrimination accuracy recorded for each *n*. Discrimination accuracy is expected to rise as more voxels are included reaching a maximum when all voxels carrying a discriminable signal are used. After this point, adding voxels will decrease discrimination accuracy because these voxels just add noise (e.g. the voxels corresponding to the foveal region which our stimuli did not stimulate except with a fixation cross). However, there is no *a priori* way of knowing how many voxels in a given region of interest (ROI) of a given participant will carry discriminable signal (Figure 7.4). In this experiment, in common with previous studies, I have used the value of prediction reached at 100 voxels across the conditions and visual areas.

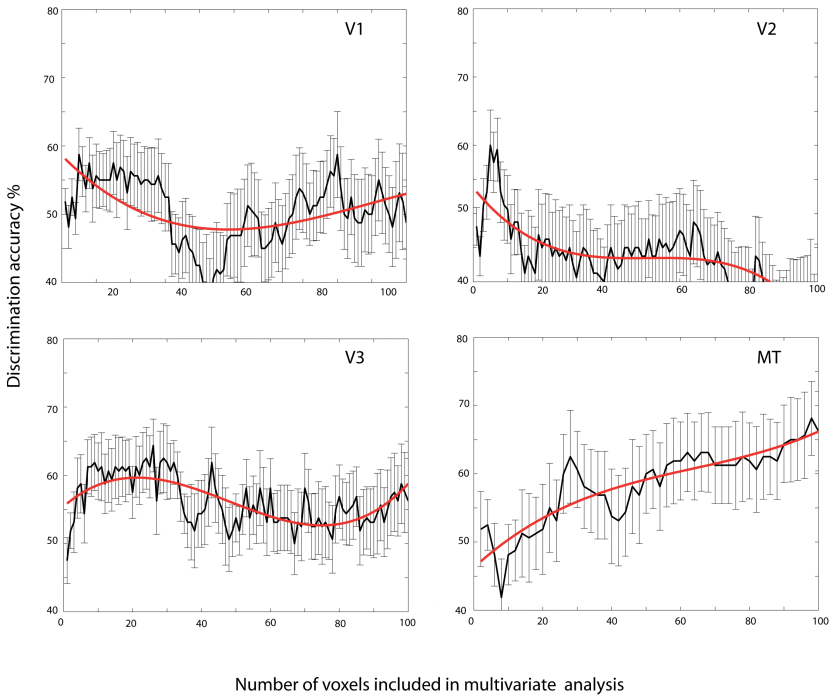


**Figure 7.4. Discrimination accuracy**

Discrimination accuracy (%) is shown as a function of the number of voxels included in the multivariate analysis for a single subject in each visual area (see Methods). Black lines represent the mean and standard error of the 16 allocations of training and test data. Red lines show a 3rd order polynomial fit. There is no *a priori* prediction for the number of useful voxels, which will depend on the functional architecture in each participant. Thus the curves of discrimination accuracy vs number of voxels are expected to rise and then fall again when voxels carrying higher signal-to-noise ratio are added (the fall may occur early or beyond 100 voxels).

## 7.4 Results

### *7.4.1 Behavioural*

The responses of two subjects indicated no consistent effect of auditory timing on visual motion during the scanning sessions, so these subjects were not included in the analysis of the auditory timing conditions. On-line behavioural data were summarized according to the average proportion of 'rightwards' responses made while viewing each condition. Results are shown in Figure 7.5, with Auditory Timing and Visual Timing conditions shown in left and right pairs of data points respectively (N=8,

visual; N =6, auditory), and within each pair of data points for conditions predicted to result in dominant leftwards versus rightwards apparent motion.

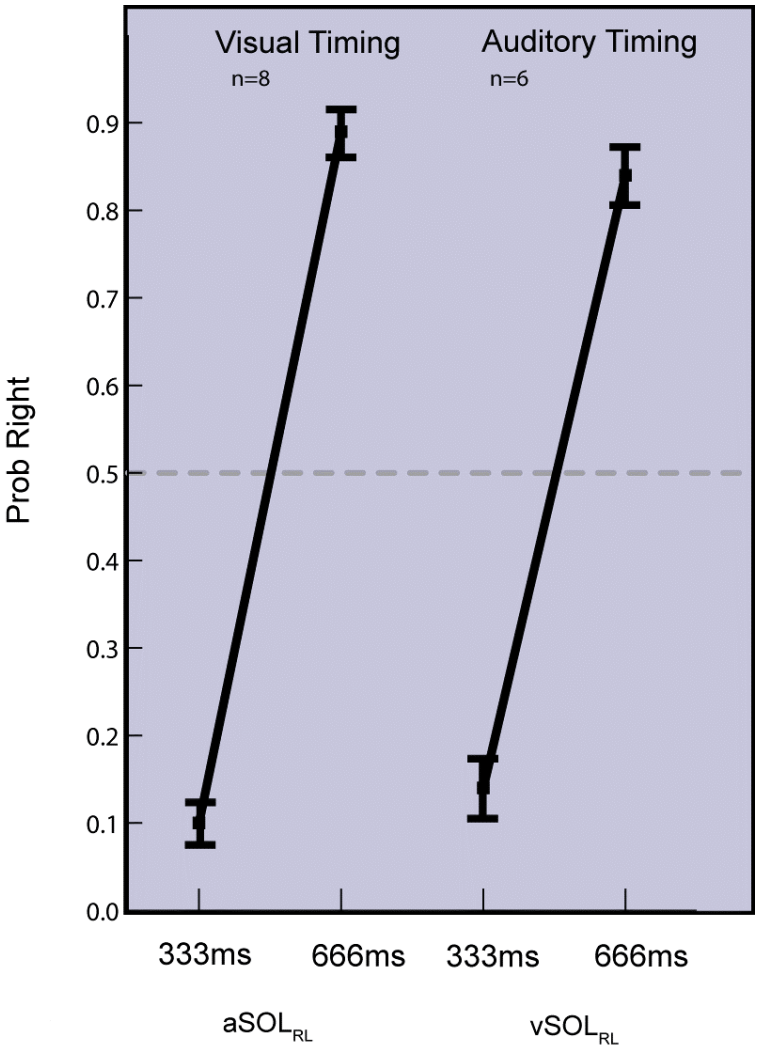


**Figure 7.5. Behavioural results.**

Results are summarized as means across 8 subjects with errorbars indicating one unit of standard error of the means. The graph plots the proportion of 'rightwards' responses, as a function of auditory or visual $SOA_{RL}$ (left and right pairs of data points respectively).

In the visual trials subjects reported 'rightwards' motion more than 'leftwards' in visual timing sequences with stimulus onset asynchrony (SOA) between R and L bars longer than for the return L-to-R direction, with 'leftwards' reports dominating for the

opposite case. The difference in responses between these SOA conditions was shown to be statistically significant [vSOArl Rightwards Response (RR) 89% SE 2%, vSOAlr RR 10% SE 1 %, paired t test t(7) = 30.3, p = <0.0001]. In the Auditory Timing trials the visual signal itself was ambiguous because the stimulus onset asynchrony was identical for the L-R and the R-L direction. However, in these trials the subjects reported significantly more 'rightwards' motion than 'leftwards' in trials where the auditory stimulus lagged the R visual bar and preceded the L visual bar [aSOArl Rightwards Response (RR) 84% SE 3%, aSOAlr RR 8% SE 4 %, paired t test t(5) = 14, p = <0.0001] The effects of auditory versus visual timing were remarkably similar [There is no statistical difference between vSOArl RR = 89% SE = 2% and aSOArl RR = 84%, SE = 3%, t(5) = 1.7, p = 0.14 or between vSOAlr RR = 10% SE = 1% and aSOAlr RR = 14%, SE = 3%, t(5) = 2.0, p = 0.10

### *7.4.2 fMRI*

The discrimination accuracy for direction of visual apparent motion was statistically above 50 % chance in visual areas V3 (mean 58%, SE 1 % t(7) = 8.9, p <.01) and MT+ (mean 58%, SE 2%,t(7) = 4.9, p <.01). The discrimination accuracy for direction of visual apparent motion did not reach statistical significance in V1 (mean 55%, SE 3%, t(7) =1.5, p = .17) and V2 (mean 54%, SE 2% t(7) = 2.1, p = .08)(Figure 7.6). Remarkably, discrimination accuracy for audiovisual apparent motion was also above chance in V3 (mean 57%, SE 1% t(5) = 4.5, p =0.01) and MT+ (mean 58%, SE 2% t(5) = 4.8, p<.01). It is important to note that in audiovisual apparent motion the visual stimulus was ambiguous. The discrimination accuracy for audiovisual apparent motion did not reach statistical significance in V1 (mean = 55%, SE 2% t(5) = 2.3, p = .07) or V2 (mean = 56%, SE 2%, t(5) = 2.5, p = .05) (Figure 7.6).

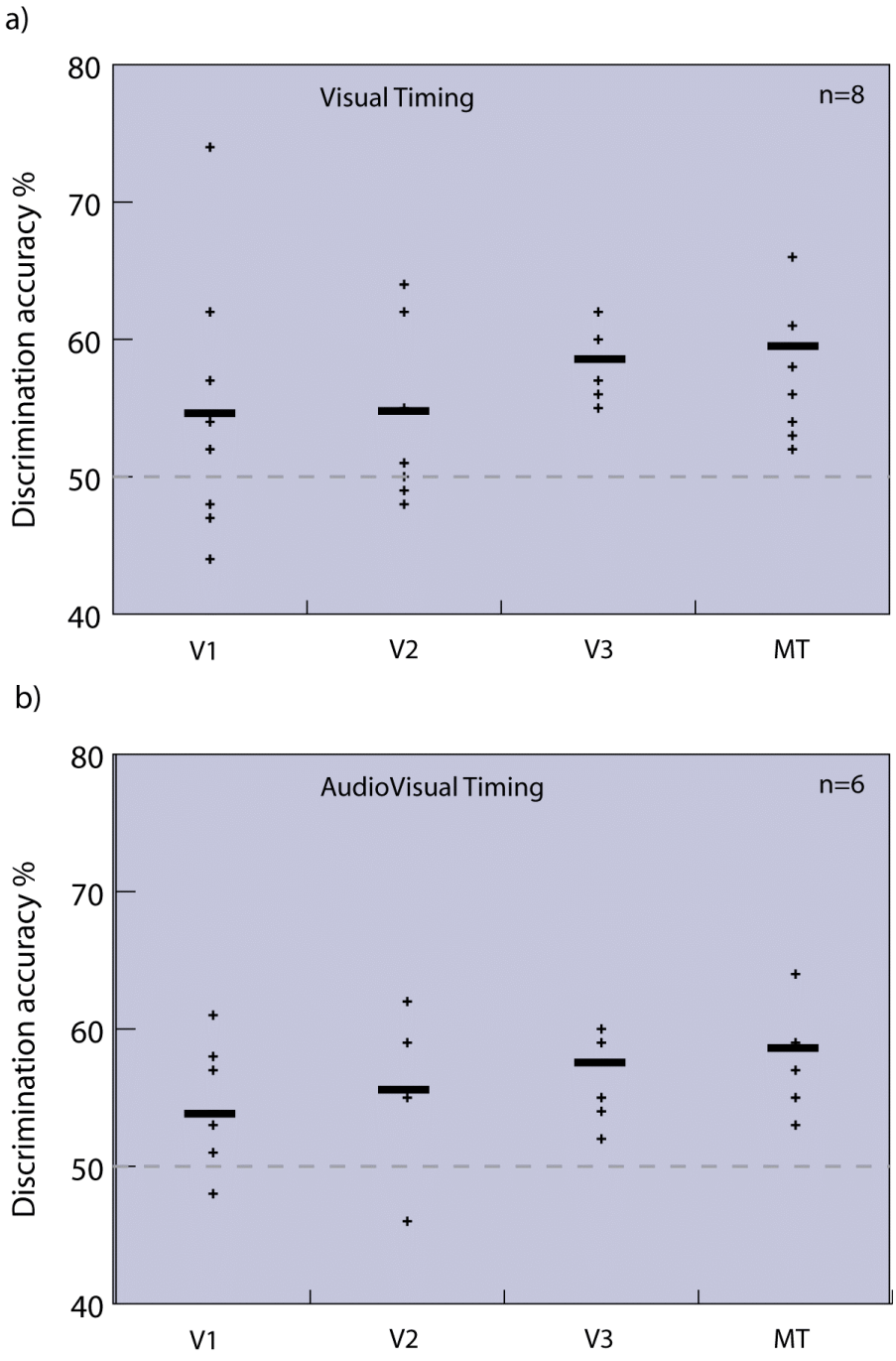


**Figure 7.6**. **Classification results**

**A.** Discrimination results for the direction of visual apparent motion based on patterns of fMRI signal across 100 voxels in each of V1, V2,V3 and MT+. The solid bar shows mean discrimination accuracy across subjects. The cross symbols represent mean accuracy for individual subjects. **B**. Discrimination results for the Audio timing condition, based on patterns of fMRI signal across 100 voxels in each of V1, V2, V3 and MT+ (symbols as for A).

## 7.5 Discussion

Recent fMRI studies have used multivariate pattern classification methods to demonstrate that it is possible to decode the direction of visual motion that is perceived from the activity in early visual areas and V5/MT+ (Kamitani and Tong, 2006). This decoding depends on the assumption that that each voxel may have a weak but true bias in direction selectivity, and therefore an ensemble of many such voxels might provide distinct patterns of activity for different perceived motion directions, which may then be discriminated using a pattern-classification algorithm (Kamitani and Tong, 2006). The present study demonstrates this for the first time with unimodal long-range apparent motion, finding reliable accuracy for leftwards and rightwards apparent motion (as determined by the timing of the visual stimuli), in cortical areas V3 and MT+. Though a new finding in itself, the involvement of such areas is perhaps not surprising given that they are also activated in univariate comparisons of motion versus static (Tootell et al., 1995;Zeki et al., 1991), and show direction-specific adaptation (Huk and Heeger, 2002;Seiffert et al., 2003). More surprising is that I also demonstrate comparable discrimination accuracy in the same areas when motion direction is determined purely by the timing of accompanying auditory stimuli, rather than by the timing of the visual events as in the traditional unimodal case above. These physiological results substantiate the claim, from earlier behavioural studies, that auditory timing signals can modulate traditionally unimodal visual motion processing, and just as effectively as purely visual timing signals (Freeman and Driver, 2008).

Traditionally, it has been assumed that multisensory integration occurs after sensory signals have undergone extensive processing in unisensory cortical regions. However, recent studies in monkeys and humans show multisensory convergence at low-level stages of cortical sensory processing previously thought to be exclusively unisensory, for a review see Foxe and Schroeder (2005). These results for audio-visual apparent motion provide additional support for such convergence, showing that auditory timing signals can influence activity in visual motion areas, though for the first time generalizing to long-range apparent motion and the influence of auditory timing.

The present findings compliment the results in previous chapters in which I demonstrated that responses in early human visual cortex can be altered by sound, and that they reflect subjective perception rather than the physically present visual stimulus. However the actual phenomenon of auditory-driven apparent motion (see also (Freeman and Driver, 2008) contrasts with the previous chapters (Watkins S et al., 2007;Watkins et al., 2006) and its antecedents (Shams et al., 2000): here subtle differences in auditory timing cause a switch between opposite directions of perceived visual motion; conversely in the past studies the illusory presence of an additional visual flash depends on the gross presence or absence of an additional sound. The fMRI findings from the present and past studies also differ markedly. In the earlier study, a correlate of illusory flashes was found in V1, while responses in V2 and V3 though similar did not reach statistical significance (Watkins S et al., 2007;Watkins et al., 2006). Here, we find differences in activity evoked by leftwards versus rightwards motion that were significant in V3 and MT+, but not in V1 and V2, despite a similar weak trend. It is interesting to consider whether the difference in the effect of auditory stimulation on V3 versus V1 between studies may be related to the above described phenomenological differences, where the sound can induce either a switch in perceived motion (as here) or a gross visual hallucination, as in previous studies (Watkins S et al., 2007;Watkins et al., 2006).

These results do not define how these auditory influences on visual processing may arise. However they are consistent with recent anatomical research where retrograde tracer injections were used to identify projections from primary auditory cortex, auditory association areas and the superior temporal polysensory area (STP) to the areas of V1, V2 and V3 representing the peripheral visual field (Clavagnier et al., 2004;Falchier et al., 2002;Rockland and Ojima, 2003). These connections have a laminar signature and a termination pattern consistent with feedback or lateral type connections(Rockland and Ojima, 2003). These connections are much more numerous in V2 and V3 than in V1 and may mediate the transfer of auditory information to the visual system.

## 7.6 Conclusion

The present results indicate that auditory timing can alter the perceived direction of visual apparent motion. I also demonstrate, for the first time, that the direction of long range visual apparent motion can be predicted from cortical activity in V3 and MT+. Remarkable the cortical activity in V3 and MT+ can still be used to predict the direction of audiovisual apparent motion.

# Chapter 8: Multimodal signals in primary visual cortex

## 8.1 Introduction

The integration of information gathered from different senses is vital to a coherent perception of the world. In the past it has often been assumed that such integration occurs in 'higher cortical areas 'after sensory signals have undergone extensive processing in unisensory cortices. However, recent studies in monkey and humans have suggested multisensory convergence at low-level stages of cortical sensory processing previously thought to be exclusively unisensory (for a review see Foxe and Schroeder, 2005). This increasing evidence for multisensory integration has led to the proposal of several different explanatory accounts of these effects.

### *8.1.1 The multisensory neocortex.*

This account postulates that most (or perhaps all) of the neocortex is essentially multisensory (or at least contains some multisensory neurons). Recent studies have demonstrated direct cortico-cortical routes between sensory areas including monosynaptic connections between primary auditory cortex and primary visual cortex in macaque (Clavagnier et al., 2004;Falchier et al., 2002) and ferret (Bizley et al., 2007).

These connections may link primary sensory cortices without the involvement of higher order multisensory areas. However, the data thus far suggest that these connections are relatively sparse and their function unclear. It is also possible that they are involved in relatively non specific modulations (i.e. arousal or alerting). Human event-related potential work has demonstrated interactions between auditory and visual (Fort et al., 2002;Giard and Peronnet, 1999;Molholm et al., 2002) or somatosensory (Foxe et al., 2002;Murray et al., 2005) stimuli at very short-latency (~46ms). These demonstrations of temporally early modulation of unisensory cortices by multisensory stimulation might conceivably reflect such direct connections (Felleman and Van Essen, 1991;Schroeder and Foxe, 2005).

### *8.1.2 Feedback from higher cortical areas*

A further possible account for multisensory influences in sensory specific cortices is that they reflect feedback influences from higher multisensory convergence zones. This perspective would retain the basic description of sensory specific cortices defined on the basis of their feedforward inputs. Several studies in macques have reported a relatively late modulation of A1 activity due to visual costimulation (Bizley et al., 2007;Ghazanfar et al., 2005). A more recent fMRI study in humans (Noesselt et al, 2007) revealed that correspondence between temporal patterning of auditory and visual streams may induce feedback from STS onto primary visual and auditory cortex.

These accounts have often been considered as rival views but it is important to state that they are not mutually exclusive. They may well coexist and different processes may be more important for different multisensory effects.

### *8.1.3 Multisensory influences on primary visual cortex*

Previous research on audiovisual interaction has tended to concentrate on two main areas. The first of which is a comparison between an audiovisual stimulus and the sum of an auditory and visual stimulus presented alone. These types of studies are vulnerable to alerting and attentional confounds. When a visual stimulus is accompanied by an auditory stimulus any change in visual processing (comparing the audiovisual stimulus to the visual stimulus alone) could conceivable be explained by the subjects paying more attention to the visual stimulus (visual attention is well know to increase cortical activity). The second type of study typically examines a multisensory 'illusion' where a change in visual perception is caused by a concurrent auditory stimulus. In these studies there is certainly an influence of the auditory stimulus on processing in primary visual cortex. However this change in activity could represent feedback from higher cortical areas. It is well know that mental imagery and perceptual decisions can affect processing in V1 (Ress and Heeger, 2003).

In chapters 5, 6 and 7 I demonstrated that activity in early visual cortex follows multisensory perception rather than the physically present visual stimulus. As I have previously discussed there is evidence that visual evoked potentials and fields are modified at short latency in association with the illusion (Bhattacharya et al., 2002;Shams et al., 2001;Shams et al., 2005), raising the possibility that audio-visual interactions responsible for illusory perception might occur in retinotopic visual cortices. However, it remains possible that these effects result from feedback from higher cortical areas rather than direct crossmodal processing in primary visual areas.

In this chapter I seek to explore these issues further. I present work investigating whether local spatial patterns of cortical activity in early visual cortex could distinguish between two audiovisual objects. Subjects were presented with a visual stimulus that could be accompanied by one of two different auditory stimuli. The auditory stimuli had no behavioural relevance or effect on visual perception. However on the basis of their multisensory properties they would potentially combine with the visual stimulus represent different multisensory objects.

If there is an early, low-level, anatomical connection between the auditory and visual cortices one might predict that a different auditory stimulus accompanying a visual stimulus would subtly alter visual processing. However, if this process required a change in audiovisual perception and a post perceptual feedback loop there should be no change in the activity in visual cortex.

I used high field BOLD fMRI to investigate whether human retinotopic areas V1-V3 contained multisensory information. Given that visual perception was postulated to be unchanged by the subtle change in the auditory stimulus I predicted that the fMRI signals analysed in a univariate (conventional, voxel-by-voxel) fashion effect would not be affected by the addition of an auditory signal to a visual stimulus. Therefore in this study I utilized recently developed multivariate pattern recognition techniques to examine the pattern of activity over V1-V3.

## 8.2 Methods

### *8.2.1 Subjects*

Eight young adults (2 females, 18-35 years old, right handed) with normal hearing and normal or corrected to normal vision gave written informed consent to participate in the study, which was approved by the local ethics committee.

### *8.2.2 Stimuli*

Visual stimuli were projected from an LCD projector (NEC LT158, refresh rate 60 Hz) onto a circular projection screen at the rear of the scanner. The subjects viewed the screen via a mirror positioned within the head coil. The auditory stimuli were presented binaurally using electrostatic headphones (KOSS, Milwaukee, USA. Model: ESP 950 Medical) custom adapted for use in the scanner. All stimuli were presented using MATLAB (Mathworks Inc.) and COGENT 2000 toolbox (www.vislab.ucl.ac.uk/cogent/index.html). Visual stimuli were tilted gratings (spatial frequency 0.8cpd and Michelson contrast 90%) presented within a smoothed annular window that subtended from 4° to 8° eccentricity. Visual stimuli were presented in 16 second blocks during which grating with one of two possible orientations (either -45°($V_L$) or +45°( $V_R$)) were presented and continuously contrast reversed with a frequency of 4 hz. The auditory stimuli consisted of two sine waves with frequencies of 400Hz ($A_1$) and 450Hz ($A_2$), duration 16s (with a ramp time of 0.5s at each end of the sound wave envelope) and volume 95dB. The sound intensity (SPL) produced by the headphones was measured while the headphones were a suitable distance away from the scanner using a sound meter (Radioshack 33-2055).

### *8.2.3 Procedure*

On each experimental trial, participants were presented with a visual stimulus either alone or accompanied by an auditory stimulus. Participants viewed the visual stimulus (grating) passively while monitoring a change in the central fixation spot. There were two control conditions where either no stimulus was presented or just an auditory stimulus and a fixation spot were presented. In total there were nine different trial types ($V_L$ $A_0$, $V_L$ $A_1$, $V_L$ $A_2$, $V_R$ $A_0$, $V_R$ $A_1$, $V_R$ $A_2$, $A_1$, $A_2$ & null). Each run consisted of 18 trials of 13 seconds. The interval between trials was 3 seconds. Participants maintained central fixation throughout. Eye position data was collected

during the trials to ensure participants maintained fixation. Each participant completed 5 runs of 18 trials making a total of 90 trials divided equally between the different trial types. Trials were pseudo-randomly distributed within a run.

### *8.2.4 fMRI data acquisition*

A 3T Siemens Allegra system was used to acquire both T1 weighted anatomical images and T2*-weighted echoplanar (EPI) images with Blood Oxygenation Level Dependent contrast (BOLD). Each EPI image volume comprised of forty 3mm axial slices with an in-plane resolution of 3x3mm positioned to cover the whole brain. Data was acquired in five runs, each consisting of 120 volumes. The first five volumes of each run were discarded to allow for T1 equilibration effects. Volumes were acquired continuously with a TR of 2.6s per volume. During scanning, eye position and pupil diameter were continually sampled at 60Hz using long-range infrared video-oculography (ASL 504LRO Eye Tracking System, Mass). Eye movements were monitored on-line via a video screen for all subjects.

### *8.2.5 Data analysis*

Eye tracking data were analysed with MATLAB (Mathworks Inc., Sherborn MA). Blinks and periods of signal loss were removed from the data. Mean eye position, expressed as a distance from fixation, was then computed for each trial type and every participant from whom data were available. A repeated-measures ANOVA was used to establish whether mean eye position deviated significantly from fixation, or between conditions.

#### *8.2.5.1 fMRI preprocessing*

Functional imaging data were analyzed using Statistical Parametric Mapping software (SPM2, Wellcome Department of Imaging Neuroscience, University College London). All image volumes were realigned spatially to the first image. Resulting image volumes were coregistered to each subject's structural scan. Activated voxels in each experimental condition for each subject were identified using a statistical model containing boxcar waveforms representing each of the experimental conditions,

convolved with a canonical hemodynamic response function and mean corrected. Multiple linear regression was then used to generate parameter estimates for each regressor at every voxel for every subject. Data were scaled to the global mean of the time series and high pass filtered (cut-off: 0.0083 Hz) to remove low-frequency signal drifts.

#### *8.2.5.2 Retinotopic analyses*

To identify the boundaries of primary visual cortex, standard retinotopic mapping procedures were used (Sereno et al., 1995;Teo et al., 1997;Wandell et al., 2000). Flashing checkerboard patterns covering either the horizontal or vertical meridian were alternated with rest periods for 16 epochs of 26 s over a scanning run lasting 165 volumes. SPM2 was used to generate activation maps for the horizontal and vertical meridians. Mask volumes for each region of interest (left and right V1, V2d, V2v, V3d, V3v) were obtained by delineating the borders between visual areas using activation patterns from the meridian localisers. We followed standard definitions of V1 together with segmentation and cortical flattening in MrGray (Teo et al., 1997;Wandell et al., 2000).

#### *8.2.5.3 Functional masks*

To identify visually responsive voxels to be used in the pattern classification, in appropriate retinotopic regions for our annulus stimuli, functional masks were obtained for each kind of stimulus. Using the mask volumes for left and right V1, V2 and V3, we identified a 100 voxels that showed the strongest activation for the comparison of all trials on which visual stimulation was present compared to null events, employing the regression analysis described above. This comparison identified voxels activated by the annular visual stimulus in each of the retinotopic areas determined by the independent meridian mapping procedure.

Having thus independently identified the stimulus representation in V1-V3 the first analytic step was to extract and average regression parameters resulting from the analysis of the main experimental time-series (described above). This procedure reliably yielded estimates of percentage signal change for each condition averaged

across voxels in V1, V2 and V3 that responded to the visual stimulus for every participant (Figure 8.2.).

### *8.2.5.4 Pattern classification*

Based on the protocols for multivariate analysis outlined previously (Haynes and Rees, 2005) and in chapter 7, I extracted, for each participant, the raw BOLD signal from 100 voxels in each of V1, V2 and V3 representing the visual stimulus (Figure 8.1). The voxels were selected using the functional masks described above, so that they came from appropriate retinotopic regions responsive to the annular stimulus. One part of the data was used to train a multivariate pattern classifier to distinguish between the activity patterns for each orientation. Based on this training, the classifier then attempted to blindly classify the orientations represented in the other, independent, part of the data (the "test"). This was repeated for different allocations of the data to training and test, and mean classification accuracy was obtained for that stimulus for that area (V1, V2 or V3) for that participant.

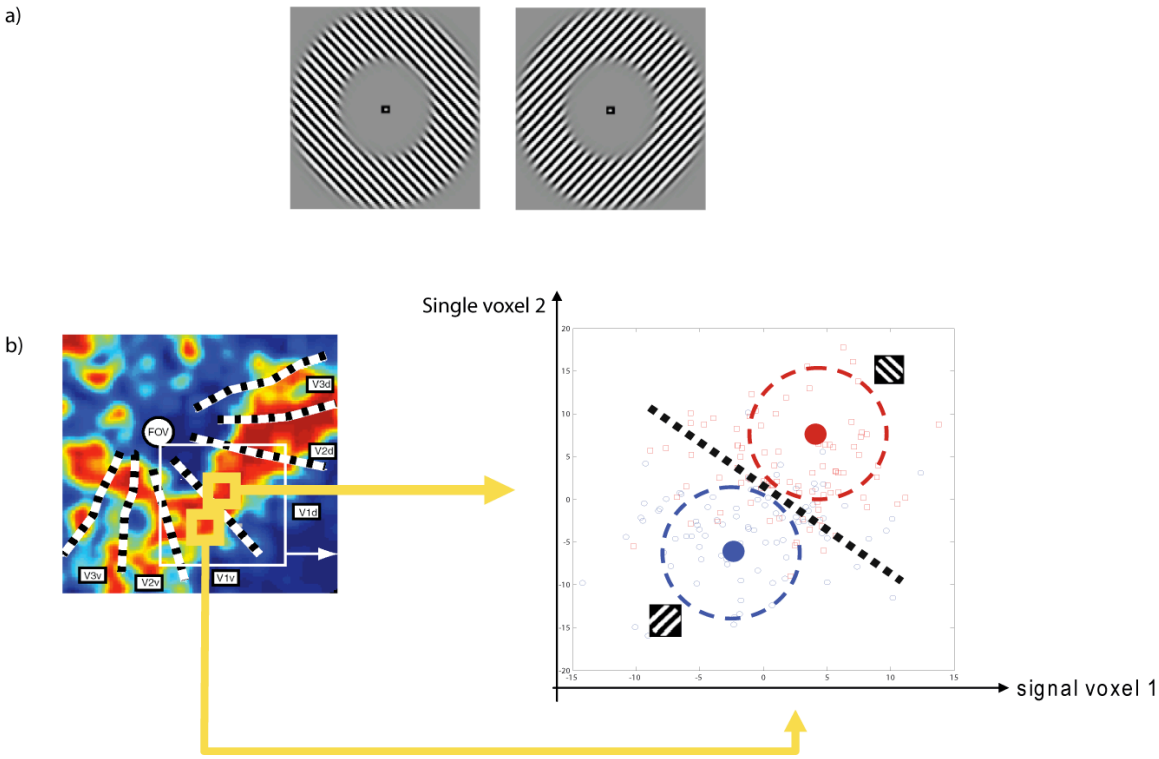


**Figure 8.1 Orientation selectivity of fMRI responses.**

a) The visual stimuli used for this experiment were contrast–reversing grating with orthogonal orientation. b) All voxels activated by the visual stimulus regardless of the accompanying auditory stimulus. The signal from each individual voxel is not sufficient to reliably discriminate between the orientations of the gratings because the response distributions are largely overlapping. However the response distributions can be separated by taking into account the combination of responses in two or more voxels (adapted from Haynes and Rees, 2006).

A total of 90 blocks of data were acquired for each subject, divided equally between the nine conditions. Each block of data made up one trial and contained 5 scans.

#### *8.2.5.4.1 Orientation discrimination*

Multivariate pattern classification was used to predict orientation for each type of stimulus from distributed response patterns in the V1-V3 ROIs, following the method outlined previously (Haynes and Rees, 2005). i.e. Comparing V-left trials ($V_L$ $A_0$, $V_L$ $A_1$ and $V_L$ $A_2$) to V-right trials ($V_R$ $A_0$, $V_R$ $A_1$ and $V_R$ $A_2$) The 5 images from 29 of the blocks for each orientation were assigned to a *training* data set (a total of 290 volumes), while the remaining 5 images in the 30$^{th}$ block were assigned to be the *test* data set. Classification performance was assessed using linear discriminant analysis with m-fold cross validation (Duda et al., 2001). Note that training and test datasets were from independent blocks, and that I used the raw fMRI signal in the coregistered images (the general linear model was used only for identifying the voxels to be included, as described above). This cycle was repeated 30 times using different blocks for training and test, and mean accuracy taken over these 30 cycles.

#### *8.2.5.4.2 Crossmodal classification*

To discriminate the presence versus absence of an auditory stimulus, I compared the trials with visual stimuli alone ($V_L$ $A_0$ and $V_R$ $A_0$) to the audiovisual trials ($V_L$ $A_1$ and $V_R A_1$). To discriminate the frequency of the auditory stimulus accompanying the visual stimulus I compared ($V_L$ $A_1$ and $V_R$ $A_1$) with ($V_L$ $A_2$ and $V_R$ $A_2$). In both cases 5 images from 19 of the blocks for each condition were assigned to the training set and the remaining 5 images of the 20$^{th}$ block to the test data set. As above the

classification performance was assessed using linear discriminant analysis with m-fold cross validation. This cycle was repeated 20 times using different blocks for training and test, and mean accuracy taken over these 20 cycles.

On each cycle, all voxels representing the stimulus were first rank ordered according to their t-value for the difference between the relevant contrast. Then the *n* voxels with the highest t values were entered into the classification algorithm; this was repeated for a range of *n* (see below), with discrimination accuracy recorded for each *n*. Discrimination accuracy is expected to rise as more voxels are included reaching a maximum when all voxels carrying a discriminable signal are used. After this point, adding voxels will decrease discrimination accuracy because these voxels just add noise (e.g. the voxels corresponding to the foveal region which our stimuli did not stimulate except with a fixation cross). However, there is no *a priori* way of knowing how many voxels in a given region of interest (ROI) of a given participant will carry discriminable signal. It is clearly not appropriate to pick the number of voxels with the maximum prediction because this would introduce a type 1 error into the results. In this experiment, in common with previous studies, I have used the value of prediction reached at 100 voxels across the conditions and visual areas.

### *8.3 Results*

#### *8.3.1 Eye position data*

Participants were requested to maintain fixation at the centre of the display. During scanning eye position was monitored on-line in all participants to ensure participants successfully maintained fixation throughout the experiment sessions. A repeated-measures ANOVA showed no statistical difference in mean eye position from fixation, or between conditions for the eight participants ($F(7,49) = .89$, $p = .24$).

#### *8.3.2 Functional MRI*

##### *8.3.2.1 Univariate analyses*

Initially a two-way within subjects ANOVA on the data from each visual area was conducted. As the stimulus was circularly symmetric and auditory stimuli presented binaurally measurements were combined across hemispheres to produce a single averaged measure for V1, V2 and V3. The factors were visual stimulus (null, left tilted grating ($V_L$), right tilted grating ($V_R$)) and auditory stimulus (absent ($A_0$), low frequency tone ($A_1$), and high frequency tone ($A_2$)). As expected there was a strongly significant main effect of visual stimulus (V1:[F(2,16)=331, p<.0001] ; V2:[F(2,16)=146, p<.0001]; V3:[F(2,16)=414, p<.0001] since responses were obviously greater when a visual stimulus was present rather than absent. There were no statistically significant differences in the mean activity evoked comparing a leftward and rightward visual grating in any retinotopic visual area (V1:[t(7)=1.33, p<.0.9] ; V2:[t(7)=1.07, p = 0.9]; V3:[t(7)=0.7, p=0.5]. There was no main effect of the auditory stimulus on the mean activity in any retinotopic visual area (V1:[F(2,14)=1.1, p=0.37] ; V2:[F(2,14)=0.285, p<.75]; V3:[F(2,14)=2.6, p<.11] (see Figure 8.2).

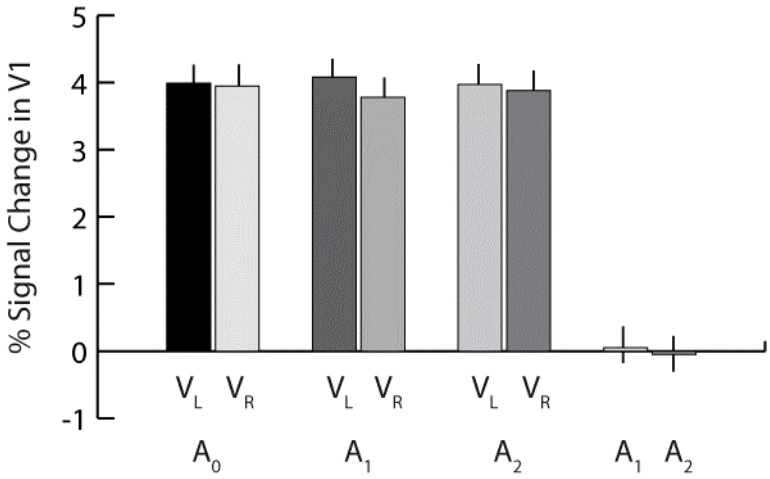


**Figure 8.2 Univariate effects in primary visual cortex.**

a) The mean percentage signal change in retinotopically defined V1 (see Methods) is shown for the conditions $V_L$ (visual grating orientated $45^0$ to the left) and $V_R$ (visual grating orientated $45^0$ to the right) accompanied by $A_0$ (no auditory stimulus), $A_1$ (high frequency tone) or $A_2$ (low frequency tone). Data shown are averaged across the eight subjects (see Methods for further details) with error bars representing the standard error of the mean,

### *8.3.2.2 Multivariate analyses*

Discrimination accuracies for the orientation of the grating (left or right tilted) were significantly greater than chance (50%) in visual areas V1 (mean 58%, t(7) = 8.9, p <.01), V2 (mean 58%, t(7) = 8.9, p <.01), and V3 (mean 58%, t(7) = 8.9, p <.01) (see Figure 8.3)

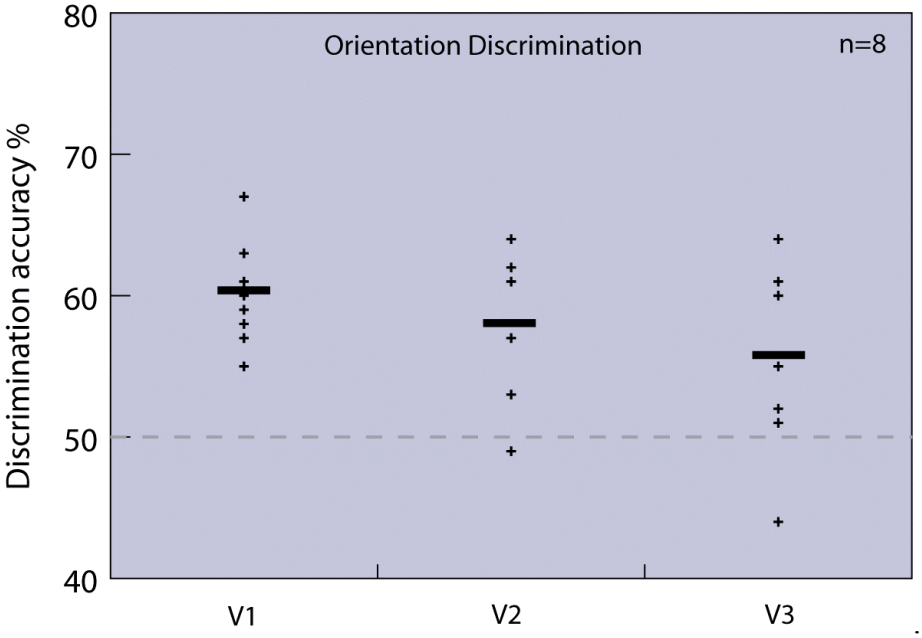


**Figure 8.3. Accuracy of prediction of orientation from visually active voxels in V1-V3.**

Discrimination results for the orientations of the visual grating based on patterns of fMRI signal across 100 voxels in each of V1, V2 and V3. The solid bar shows mean discrimination accuracy across participants. The cross symbols represent mean accuracy for individual participants.

The discrimination accuracy for presence versus absence of auditory stimuli was statistical significant in V1 (mean 55%, t(7) =2.5, p = .04) and V3 (mean 61%, t(7) = 4.9, p = .002) but was not significant in V2 (mean 54%, t(7) =1.5, p = .12) (Figure 8.4).

Discrimination accuracy for the frequency of the sound accompanying a visual stimulus was not above chance in V1 (mean 53%, t(7) =1.1, p = .29), V2 (mean 50%, t(7) =0.5, p = .96) or V3 (mean 55%, t(7) =0.37, p = .97) (see Figure 8.4).

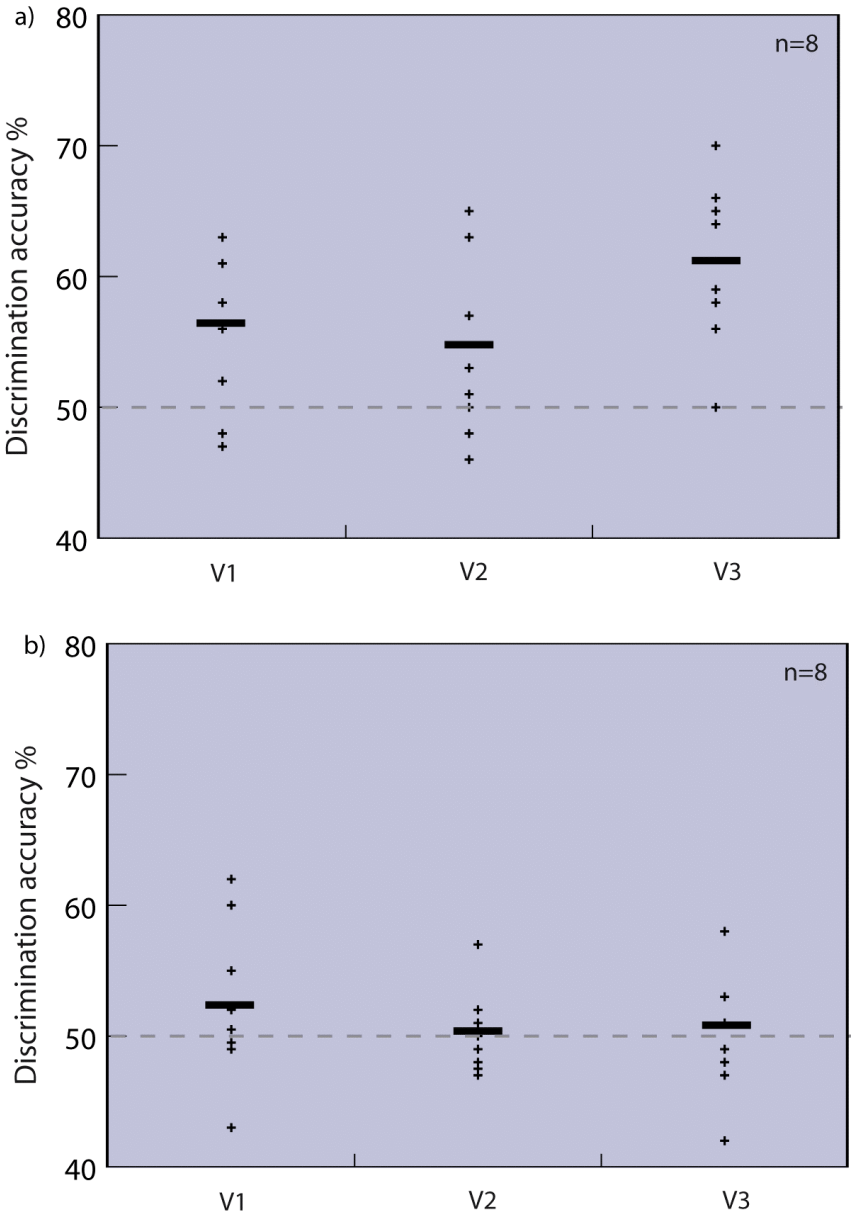


**Figure 8.4. Prediction accuracy of audiovisual stimulus from visually active voxels in V1-V3.**

a) Discrimination results for the presence versus absence of auditory stimulus based on patterns of fMRI signal across 100 voxels in each of V1, V2, & V3. The solid bar shows mean discrimination accuracy across participants. The cross symbols represent mean accuracy for individual participants. b) Discrimination results for the different audiovisual stimulus, based on patterns of fMRI signal across 100 voxels in each of V1, V2 & V3.

## 8.4 Discussion

In this chapter I confirmed that the orientation of a visual grating can reliably be predicted from cortical activity in early visual areas (Haynes and Rees, 2006).

In addition, the cortical processing in V1 & V3 was altered by the presence versus absence of an auditory stimulus. This is consistent with my previous work in chapters 5 & 6 in which the addition of a sound to a visual stimulus caused a change in visual processing compared to the visual stimulus presented alone (Chapter 5, Figure 5.4). These findings are also consistent with previous observations that behavioural or physiological responses to visual stimulation can be modified by sound (Bermant and Welch, 1976;Kitagawa and Ichihara, 2002;Morrell, 1972;Reisberg, 1978;Sekuler et al., 1997).

I did not any find evidence that the early visual cortical areas encode the frequency of an accompanying auditory stimulus. This experiment examines an unusual situation in that the auditory stimulus conveys no useful behavioural information. It could be argued that in most real world examples crossmodal perception may offer a behavioural benefit. However, these results suggest that the early visual cortex, as the name suggests, is primary devoted to encoding information about vision and does not represent auditory information unless it changes visual perception or visual attention.

Recently Lemus and colleagues (2010) examined a similar question in the auditory and somatosensory cortex of monkeys. In their study the authors employed the flutter discrimination task(Lemus et al., 2010). Monkeys were trained to discriminate vibrating stimuli that were either presented as a tactile sensation or an acoustic pulse train. During each trial, two stimuli of differing frequency were presented, interspaced with a short interval, and the animal had to decide which stimuli was higher frequency. Their results indicated that the neurons in S1 encode the tactile frequency. However, although a small percentage responded during the auditory flutter, importantly their responses did not vary with the flutter frequency. Similarly neurons in A1 encoded the acoustic, but not tactile frequency. They concluded that although neurons in both S1 and A1 responded to crossmodal stimuli, there was no encoding of

information contained in the alternative modality stimulus in the primary sensory area. In summary they suggested that multimodal encoding occurs outside of primary sensory cortices (Lemus et al, 2010).

There remains a significant body of evidence from previous studies that suggests multisensory processing can occur in primary sensory cortices (See Ghazanfar & Schroeder, 2006 and chapter 5, 6 & 7 for a review). In these previous studies the multisensory input was usually of behaviour or perceptual relevance. For example in chapters 5 & 6 the additional auditory stimulus caused a clear change in visual perception (one flash was perceived as two flashes when accompanied by two bleeps). In this study evidence was found that primary visual cortex can respond to auditory input. It was possible to predict whether an auditory stimulus is present or absent from the cortical activity in primary visual cortex. However, the activity in primary visual cortex did not contain enough information to predict what type of auditory stimulus accompanied the visual stimulus.

It is possible that multiple mechanisms of cross modal integration are utilized in the human brain and in different behavioural circumstances there would be a greater degree of crossmodal integration in primary sensory cortices. However in this study I demonstrated that in the situation where the auditory stimulus has no behavioural or perceptual relevance to a visual stimulus there is no processing of the feature attributes of the auditory stimulus in the visual cortex. Therefore, in this experiment I would suggest that crossmodal processing in primary sensory cortices is not automatic but depends on the 'relevance' of the crossmodal stimulus.

### 8.5 Conclusion

In summary, I have demonstrated that primary visual cortex can respond to auditory stimuli. This is consistent with many previous studies and my work in chapters 5,6 &7. However in the situation where an auditory stimulus has no behaviour or perceptual relevance to visual processing early visual areas do not encode information about the auditory stimulus. These results suggest that the primary role of the visual cortex is to encode visual information and stimuli from other modalities will only be processed when they influence visual processing.

# CHAPTER 9: GENERAL DISCUSSION

## 9.1 Introduction

The experiments outlined in this thesis demonstrate that the presence of a task irrelevant stimulus can exert a significant behavioural and perceptual effect. This general discussion will review these findings and the implications they have for our understanding of cortical perceptual processing. In addition, I will explore the extent to which these findings can shed light on other brain processes and indeed abnormal perception or behaviour following damage to the brain. Finally, I will discuss further experimental studies that would test hypotheses generated from the experimental data presented in this thesis. The experimental studies can be grouped according to whether they examine the effects of a task irrelevant stimulus on behaviour (chapters 3 & 4) or perception (chapters 5, 6, 7 & 8).

## *9.2 Attentional capture - the effect of a task irrelevant distractor on behaviour and cortical processing*

Complex scenes are often cluttered with many different stimuli. At any given time only a small fraction of the information received can be selected for further processing. In real life, people are usually able to focus on stimuli relevant for the task at hand. This can be achieved by using knowledge and expectations to focus attention on task relevant signals rather than competing irrelevant stimuli. Despite this top-down control, a unique stimulus can 'capture' attention, even when task-irrelevant. Although distracting subjects from their current task, such attentional capture may have a survival advantage, as a unique stimulus may often convey important information about the environment.

### *9.2.1 Auditory attentional capture*

In chapter 3 I utilized a behavioural paradigm that allowed me to distinguish cortical responses that were related to increased auditory variation in the stimulus (i.e. the presence of an irrelevant feature singleton regardless of whether it captures attention)

to be distinguished from cortical responses that were specific to auditory attentional capture (i.e. those related to presence of a distractor feature singleton that caused behaviourally defined attentional capture). The findings suggested that a ventral network, involving bilateral superior temporal gyri and the left inferior frontal gyrus, responds to auditory variability regardless of its behavioural consequence. In contrast, activation of a more dorsal network comprising of left precentral gyrus, right superior parietal gyrus and right intraparietal sulcus responds specifically to capture of attention by an auditory feature singleton.

The cortical response to auditory variability demonstrated in this study is similar to the network of areas that are proposed to mediate the mismatched negativity (MMN). The MMN is a cortical response to an auditory oddball or deviant stimulus in a stream of repeated or familiar events. Previous studies have demonstrated that generation of the MMN has consistently been shown to involve bilateral superior temporal gyri (Deacon et al., 1998;Giard et al., 1990;Opitz et al., 1999b). However the role of the frontal cortex in MMN is less clear with some studies finding no frontal activation (Opitz et al., 1999a) and others suggesting involvement of the right inferior frontal gyrus (Opitz et al., 2002) or bilateral inferior frontal gyrui (Doeller et al., 2003;Molholm et al., 2005).

The dorsal frontoparietal network associated with auditory attentional capture is similar to that associate with visual attentional capture. Thus supporting the hypothesis that vision and audition share a common cortical attentional network. These findings are the first demonstration of the cortical network mediating attentional capture in audition.

### *9.2.2 Visual attentional capture*

The neural substrates of singleton capture in search have previously been investigated (de Fockert et al, 2004). In that earlier study, neural activity was measured via functional magnetic resonance imaging (fMRI) as subjects viewed similar search displays to Theeuwes (1991). Subjects searched for a circle among diamonds, and on 25% of the trials the target was a colour singleton, whereas on another 25% of the trials one of the distractors was a colour singleton. Although there was no measured

neural activity specifically related to the colour singleton target, the presence of a colour singleton distractor led to bilateral activation within the dorsal parietal cortex and left frontal cortex relative to when no colour singleton was present. The parietal activity was construed to reflect shifts of attention to the distractor item (e.g., Corbetta and Shulman, 2002) and frontal activity to reflect the resolution of subsequent competition between the salient distractor and the target.

An unresolved question in these attentional capture studies was the behavioural and cortical response to varying the level of salience of the distractor singleton. Previous studies suggest that if a distractor singleton is more salient than the target it will capture attention. However, the response to parametrically changing the level of salience of the distractor has not been investigated. It thus remains an open question as to whether attentional capture might reflect either an all-or-none response to the presence of a distractor, regardless of salience; or whether varying distractor salience might be associated with parametrically varying levels of attentional capture. In chapter 4 I investigated this important question. Initially, I demonstrated that a bilateral network of dorsal parietal (principally IPS and SPL) and prefrontal cortex were associated with visual attentional capture. This replicates and extends the previous study on attentional capture by de Fockert (de Fockert et al, 2004).

More importantly, I demonstrated a positive correlation between the level of salience of the distractor singleton and behavioural interference (See Chapter 4, Figures 4.2 & 4.3). Even the lowest level of distractor salience caused a significant disruption to behaviour. When the singleton feature was present in the target ('target singleton') I observed a small but significant facilitation effect. The facilitation effect did not increase as the salience of the target singleton increased (Chapter 4, Figure 4.3). There are two possible explanations for these effects. Firstly, that attentional capture only occurred on a subset of the trials and this number increased as the salience of the distractor increased. This explanation implies I should also see the facilitation effect increase as the salience of the target singleton increases. In contrast, in this study I did not demonstrate a significant change in the facilitation effect across the levels of salience. An alternative explanation is that the level of salience does not affect the initial stimulus driven transfer of attention to the distractor singleton but effects the

speeds with which attention can be disengaged from the distractor and transferred back to the target (see Theeuwes, 2010).

The response of a relatively restricted set of cortical areas was correlated to increased distractor salience. These included areas in bilateral frontal cortex and left parietal cortex but not right parietal cortex. I would suggest that these areas are involved in resisting distractor and reorientating attention back to the target. Atttenional capture produced strong activity in the right parietal cortex which did not change with distractor salience. This may imply that right parietal activity is involved in the initial bottom up shift of attention rather than top down control of attentional capture. The suggestion that left and right parietal cortex have different roles in attentional capture is generally consistent with previous TMS studies (utilizing a wide variety of spatial attention tasks) that have shown disruption of spatial attention processes following TMS to the right PPC but not the left (cf. Rushworth and Taylor 2006 for a review). In visual search, for example, Ellison et al. (2003) showed that TMS impaired conjunction search following right parietal TMS. Also consistent with this idea that the left and right parietal lobes subserve different critical functions is a substantial body of data based on unilateral neglect (e.g.,Heilman and Valenstein 1979). This syndrome, associated with poor attentional orienting to the affected side, is more commonly found following right parietal as opposed to left lesions.

There are a number of related areas that further research could address. Firstly, a study including visual and auditory attentional capture in the same experiment could further address the important question of whether the attenional network is truly crossmodal. Such a study would permit a direct comparison of visual and auditory attentional capture, rather than the qualitative comparison between Chapters 3 & 4 presented here. Secondarily, it would be interesting to further investigate the different functions of the fronoparietal network activated by attentional capture. For example utilizing the new multivariate analysis techniques it would be possible to test whether the pre trail activity in the pre central gyrus predicted whether a stimulus would capture attention. This would provide stronger support that this cortical area has a role in resisting distraction. But perhaps the most intriguing questions for future work relate to the dynamic interplay between the cortical regions in the attentional network. It would be important to know the timing of activation of various parts of the

described attentional network and the degree to which each area is critical to attentional capture.
The recent combing of fMRI and TMS and of fMRI and EEG is starting to provide valuable insights into these issues.

Finally further study of disorders of attention in humans such as neglect and attention deficit disorders may lead to greater understanding of these disabling conditions.

In the next four chapters I have turned my attention to the affect of an irrelevant auditory stimulus on visual perception and cortical visual processing.

### 9.3 Audiovisual integration

The integration of information from multiple senses is fundamental to our perception of the world. Traditionally, it has been assumed that multisensory integration is deferred until after sensory signals have undergone extensive processing in unisensory cortical regions. However, recent work has demonstrated that purportedly primary sensory cortices can respond to stimuli from a different modality (Foxe and Schroeder, 2005).

An important but unresolved issue that may provide insight into the function of multisensory convergence concerns how such neural interactions might be reflected in conscious perception. If activity in early sensory cortices corresponds to a particular conscious experience, then modification of that activity by converging multisensory input should be related to changes in conscious experience.

In chapters 5 & 6 I sought to address this issue by measuring brain activity in human volunteers experiencing an established audio-visual illusion, in which the presence of irrelevant sounds can modify the perception of a simple visual stimulus (Shams et al., 2000). Crucially, this illusion occurs on only a proportion of trials, with veridical perception of the visual stimulus being reported on the non-illusion trials. This means it is possible to compare trials with identical auditory and visual stimulation that nevertheless had very different perceptual outcomes. The principle results demonstrated that the activity in primary visual cortex followed audiovisual

perception rather than the physically present visual stimulus. This is consistent with increasing evidence from unisenory studies that V1 activity can be more closely related to conscious visual experience rather than physical present visual stimulus (Ress and Heeger, 2003).The present findings show that such an association of V1 activity with conscious perception extends to changes in visual perception brought about by multisensory stimulation.

In chapter 7 I utilized fMRI and multivariate pattern classification to investigate another effect of a task irrelevant auditory stimulus on visual perception. In the 'temporal ventriloquism' effect, an auditory event that either leads or lags a visual event can seem to shift the visual onset backwards or forwards in time respectively (Gebhard et al., 1959;Morein-Zamir et al., 2003;Vroomen et al., 2004). Recent work (Freeman and Driver, 2008) has demonstrated that pure auditory timing can influence visual processing of spatio-temporal patterns, namely motion. A long-debated issue concerns whether such phenomena reflect feedforward convergence of multimodal timing information in higher cortical areas, or alternatively crossmodal interactions between sensory processing in early sensory cortices (Ghazanfar and Schroeder, 2006). In chapter 7 I demonstrated for the first time using multivariate analysis that the direction of visual apparent motion can be predicted from patterns of activation in motion-responsive visual areas (V3 and MT+). Remarkably, such patterns in V3 and MT+ can also be used to predict the perceived direction of ambiguous visual apparent motion whose perceived direction is biased solely on auditory information.

From the methodological viewpoint, the use of multivariate rather than mass univariate analysis requires further validation. The multivariate technique is a very sensitive method for finding differences between sets of data, but the physiological basis and functional significance of these differences is not yet clear. Further work to clarify the underlying basis for this relatively new form of fMRI analysis is needed.

In summary in these experiments subjects were unable to ignore a task irrelevant auditory stimuli. These results suggest, at least in this experiment setup, that audiovisual integration may occur automatically. Multisensory integration has often been described as occurring automatically. Early studies investigating the response properties of single neurons in anaesthestised animals demonstrated multisensory

integration provided there was spatial and temporal concordance between the stimuli (Stein et al., 2004;Stein and Arigbede, 1972;Wallace et al., 1998). Behavioural work in humans has demonstrated that crossmodal model integration can occur preattentively (Driver, 1996;Van der Burg et al., 2008). Chapter 5 was the first study to demonstrate multisensory integration in retinotopically defined V1. These data do not precisely define how the association of visual cortex activation with illusory visual perception occurs, nor whether the modulation of visual cortex activity I observed plays a causal role in the generation of these effects. Primary visual cortex receives projections from at least twelve areas belonging to the visual cortex (Felleman and Van Essen, 1991). Recently more distant projections have been described from areas in the ventral (Distler et al., 1993b) and dorsal visual pathways and from the lateral intraparietal area (Boussaoud et al., 1990;Rockland and Van Hoesen, 1994). Several recent papers have used tracer injections to demonstrate projections from primary auditory cortex, auditory association areas and the superior temporal polysensory area (STP) to the area of primary visual cortex representing the peripheral visual field (Clavagnier et al., 2004;Falchier et al., 2002;Rockland and Van Hoesen, 1994). These connections are present in V1 but more numerus in V2 and V3. The function of these projections to early visual cortex has been the subject of much debate, they may serve to enhance perceptual capabilities; for example the addition of an auditory signal to a visual signal leads to improved detection compared to a visual signal alone (Bolognini et al., 2005;Frassinetti et al., 2002;Gondan et al., 2005;Miller, 1982;Molholm et al., 2002;Schroger and Widmann, 1998a). Thus, it is possible that these direct connections mediate the changes in visual cortical activity that I observed in chapters 5, 6 & 7.

An alternative account would be that the change in activity in early visual areas is the result of feedback form higher cortical areas after the perceptual decision has been made. In the current studies fMRI does not have the temporal resolution to distinguish between these two options. However, previous research using event-related potential recordings in human show that multisensory integration can occur very early in visual processing. For example, a change in a simple visual stimulus that is accompanied by a change in pitch of a concurrent tone can lead to modification of the ERP at very short latencies(Giard and Peronnet, 1999), and auditory clicks can modify the evoked potential to pattern stimulation in visual cortex (Arden et al., 2003). For the flash-

bleep illusion studied in chapters 5 & 6, visual evoked potentials and fields associated with the illusory perception are modified at a short latency (Bhattacharya et al., 2002;Shams et al., 2001;Shams et al., 2005) consistent with generators in early visual cortex.

In the final chapter (chapter 8) I explored the extent to which auditory information is present in primary visual cortex. In chapters 5, 6 & 7 an auditory stimulus that accompanies a visual stimulus causes a clear change in visual perception. I have clearly demonstrated that this change in perception is represented in primary visual cortex for the flash bleep illusion and V3 and MT for audiovisual apparent motion. In chapter 8 I examine the response of visual cortex to two audiovisual objects that differ by only a small change in the frequency of the auditory stimulus. This is an entirely irrelevant change in the auditory stimulus that would not be expected to induce a change in visual perception. I did not find any evidence that early visual cortical areas encode the frequency of the accompanying auditory stimulus. This experiment examines an unusual situation in that the auditory stimulus conveys no useful behavioural information. It could be argued that in most real world examples crossmodal perception may offer a behavioural benefit. However, these results suggest that the early visual cortex, as the name suggests, is primary devoted to encoding information about vision (or at least visual perception) and does not represent auditory information unless it changes visual perception or visual attention.

Taken together, these findings suggest that when an irrelevant auditory stimulus causes a change in visual perception this change can be detected experimentally in the anatomically earliest areas of visual processing, including primary visual cortex. In addition they suggest that crossmodal stimuli are not automatically processed in primary sensory areas but processing may critically depend on whether there is a change in sensory perception. I have discussed a number of different accounts and architectures that have been proposed for these newly uncovered crossmodal phenomena (chapter 8), ranging from the idea that all areas may be inherently multisensory (or perhaps less extremely, may all have at least some multisensory interneurons (Allman and Meredith, 2007) distributed among them, in differing proportions), to thalamic influences and/or direct connections between primary cortices, to the possibility that some multisensory effects may reflect feedback

influences from higher-level multisensory convergence-zones, back to otherwise sensory-specific regions.

There are a number of related questions that further research could address. Firstly, an intriguing question is the temporal and spatial limits of multisensory integration. Secondarily further investigation into the underlying cortical architecture that mediates the multisensory effect I have described.

### 9.4 Conclusion

In summary, my findings provide evidence that task irrelevant distractors can have significant effects on behaviour and brain activity. I have demonstrated the cortical areas responsible for attentional capture in audition and vision and how these areas respond when the salience of the distractor changes. In the second group of experiments I have demonstrated that irrelevant auditory stimuli can influence visual perception and processing at the anatomical earliest stages of visual processing.

Reference List

Alho,K., Escera,C., Diaz,R., Yago,E., and Serra,J.M. (1997). Effects of involuntary auditory attention on visual task performance and brain activity. Neuroreport *8*, 3233-3237.

Amedi,A., Malach,R., Hendler,T., Peled,S., and Zohary,E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. Nat.Neurosci. *4*, 324-330.

Andersen,T.S., Tiippana,K., and Sams,M. (2004). Factors influencing audiovisual fission and fusion illusions. Brain Res.Cogn Brain Res. *21*, 301-308.

Arden,G.B., Wolf,J.E., and Messiter,C. (2003). Electrical activity in visual cortex associated with combined auditory and visual stimulation in temporal sequences known to be associated with a visual illusion. Vision Res. *43*, 2469-2478.

Attwell,D. and Iadecola,C. (2002). The neural basis of functional brain imaging signals. Trends Neurosci. *25*, 621-625.

Attwell,D. and Laughlin,S.B. (2001). An energy budget for signaling in the grey matter of the brain. J.Cereb.Blood Flow Metab *21*, 1133-1145.

Bacon,W.F. and Egeth,H.E. (1994). Overriding stimulus-driven attentional capture. Percept.Psychophys. *55*, 485-496.

Barbas,H., Medalla,M., Alade,O., Suski,J., Zikopoulos,B., and Lera,P. (2005). Relationship of prefrontal connections to inhibitory systems in superior temporal areas in the rhesus monkey. Cereb.Cortex *15*, 1356-1370.

Barraclough,N.E., Xiao,D., Baker,C.I., Oram,M.W., and Perrett,D.I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J.Cogn Neurosci. *17*, 377-391.

Beauchamp,M.S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Curr.Opin.Neurobiol. *15*, 145-153.

Beauchamp,M.S., Argall,B.D., Bodurka,J., Duyn,J.H., and Martin,A. (2004). Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat.Neurosci. *7*, 1190-1192.

Beck,D.M. and Kastner,S. (2005). Stimulus context modulates competition in human extrastriate cortex. Nat.Neurosci. *8*, 1110-1116.

Benedict,R.H., Lockwood,A.H., Shucard,J.L., Shucard,D.W., Wack,D., and Murphy,B.W. (1998). Functional neuroimaging of attention in the auditory modality. Neuroreport *9*, 121-126.

Berger,T.D., Martelli,M., and Pelli,D.G. (2003). Flicker flutter: is an illusory event as good as the real thing? J Vis. *3*, 406-412.

Bermant,R.I. and Welch,R.B. (1976). Effect of degree of separation of visual-auditory stimulus and eye position upon spatial interaction of vision and audition. Percept.Mot.Skills *42*, 487-493.

Berti,S., Roeber,U., and Schroger,E. (2004). Bottom-up influences on working memory: behavioral and electrophysiological distraction varies with distractor strength. Exp.Psychol. *51*, 249-257.

Berti,S. and Schroger,E. (2004). Distraction effects in vision: behavioral and event-related potential indices. Neuroreport *15*, 665-669.

Best,V., Ozmeral,E.J., and Shinn-Cunningham,B.G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. J.Assoc.Res.Otolaryngol. *8*, 294-304.

Bhattacharya,J., Shams,L., and Shimojo,S. (2002). Sound-induced illusory flash perception: role of gamma band responses. Neuroreport *13*, 1727-1730.

Bizley,J.K., Nodal,F.R., Bajo,V.M., Nelken,I., and King,A.J. (2007). Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb.Cortex *17*, 2172-2189.

Bloch,F. (1946). Nuclear Induction. Phys.Rev 460-474.

Bolognini,N., Frassinetti,F., Serino,A., and Ladavas,E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. Exp.Brain Res. *160*, 273-282.

Boussaoud,D., Ungerleider,L.G., and 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. J Comp Neurol. *296*, 462-495.

Budinger,E., Heil,P., Hess,A., and Scheich,H. (2006). Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. Neuroscience *143*, 1065-1083.

Calvert,G.A., Bullmore,E.T., Brammer,M.J., Campbell,R., Williams,S.C., McGuire,P.K., Woodruff,P.W., Iversen,S.D., and David,A.S. (1997). Activation of auditory cortex during silent lipreading. Science *276*, 593-596.

Calvert,G.A., Campbell,R., and Brammer,M.J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr.Biol. *10*, 649-657.

Calvert,G.A., Hansen,P.C., Iversen,S.D., and Brammer,M.J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. Neuroimage *14*, 427-438.

Calvert,G.A. and Thesen,T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. J Physiol Paris *98*, 191-205.

Carrasco,M., Giordano,A.M., and McElree,B. (2006). Attention speeds processing across eccentricity: feature and conjunction searches. Vision Res. *46*, 2028-2040.

Cheal,M.L. and Gregory,M. (1997). Evidence of limited capacity and noise reduction with single-element displays in the location-cuing paradigm. J.Exp.Psychol.Hum.Percept Perform. *23*, 51-71.

Clavagnier,S., Falchier,A., and Kennedy,H. (2004). Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. Cogn Affect.Behav.Neurosci. *4*, 117-126.

Corbetta,M., Patel,G., and Shulman,G.L. (2008). The reorienting system of the human brain: from environment to theory of mind. Neuron *58*, 306-324.

Corbetta,M. and Shulman,G.L. (2002). Control of goal-directed and stimulus-driven attention in the brain. Nat.Rev.Neurosci. *3*, 201-215.

Coull,J.T., Frith,C.D., Buchel,C., and Nobre,A.C. (2000). Orienting attention in time: behavioural and neuroanatomical distinction between exogenous and endogenous shifts. Neuropsychologia *38*, 808-819.

Dalton,P. and Lavie,N. (2004). Auditory attentional capture: effects of singleton distractor sounds. J Exp.Psychol.Hum.Percept.Perform. *30*, 180-193.

de Fockert,J., Rees,G., Frith,C., and Lavie,N. (2004). Neural correlates of attentional capture in visual search. J Cogn Neurosci. *16*, 751-759.

Deacon,D., Nousak,J.M., Pilotti,M., Ritter,W., and Yang,C.M. (1998). Automatic change detection: does the auditory system use representations of individual stimulus features or gestalts? Psychophysiology *35*, 413-419.

Distler,C., Boussaoud,D., Desimone,R., and Ungerleider,L.G. (1993a). Cortical connections of inferior temporal area TEO in macaque monkeys. J Comp Neurol. *334*, 125-150.

Distler,C., Boussaoud,D., Desimone,R., and Ungerleider,L.G. (1993b). Cortical connections of inferior temporal area TEO in macaque monkeys. J Comp Neurol. *334*, 125-150.

Doeller,C.F., Opitz,B., Mecklinger,A., Krick,C., Reith,W., and Schroger,E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. Neuroimage *20*, 1270-1282.

Dosher,B.A. and Lu,Z.L. (2000). Mechanisms of perceptual attention in precuing of location. Vision Res. *40*, 1269-1292.

Dougherty,R.F., Koch,V.M., Brewer,A.A., Fischer,B., Modersitzki,J., and Wandell,B.A. (2003). Visual field representations and locations of visual areas V1/2/3 in human visual cortex. J Vis. *3*, 586-598.

Downar,J., Crawley,A.P., Mikulis,D.J., and Davis,K.D. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. J.Neurophysiol. *87*, 615-620.

Driver,J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. Nature *381*, 66-68.

Duda,O.R., Hart,P.E., and Strok D.G (2001). Pattern Classification. (New York: Wiley).

Engel,S.A., Rumelhart,D.E., Wandell,B.A., Lee,A.T., Glover,G.H., Chichilnisky,E.J., and Shadlen,M.N. (1994). fMRI of human visual cortex. Nature *369*, 525.

Eriksen,C.W. and Hoffman,J.E. (1973). The extent of processing of noise elements during selective encoding from visual displays. percept psychophs *14*, 155-160.

Escera,C., Alho,K., Winkler,I., and Naatanen,R. (1998). Neural mechanisms of involuntary attention to acoustic novelty and change. J Cogn Neurosci. *10*, 590-604.

Escera,C., Yago,E., and Alho,K. (2001). Electrical responses reveal the temporal dynamics of brain events during involuntary attention switching. Eur.J.Neurosci. *14*, 877-883.

Escera,C., Yago,E., Corral,M.J., Corbera,S., and Nunez,M.I. (2003). Attention capture by auditory significant stimuli: semantic analysis follows attention switching. Eur.J.Neurosci. *18*, 2408-2412.

Falchier,A., Clavagnier,S., Barone,P., and Kennedy,H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. J Neurosci. *22*, 5749-5759.

Felleman,D.J. and Van Essen,D.C. (1991). Distributed hierarchical processing in the primate cerebral cortex. Cereb.Cortex *1*, 1-47.

Folk,C.L., Remington,R.W., and Johnston,J.C. (1992). Involuntary covert orienting is contingent on attentional control settings. J.Exp.Psychol.Hum.Percept.Perform. *18*, 1030-1044.

Fort,A., Delpuech,C., Pernier,J., and Giard,M.H. (2002). Early auditory-visual interactions in human cortex during nonredundant target identification. Brain Res.Cogn Brain Res. *14*, 20-30.

Foxe,J.J. and Schroeder,C.E. (2005). The case for feedforward multisensory convergence during early cortical processing. Neuroreport *16*, 419-423.

Foxe,J.J., Wylie,G.R., Martinez,A., Schroeder,C.E., Javitt,D.C., Guilfoyle,D., Ritter,W., and Murray,M.M. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. J Neurophysiol. *88*, 540-543.

Frassinetti,F., Pavani,F., and Ladavas,E. (2002). Acoustical vision of neglected stimuli: interaction among spatially converging audiovisual inputs in neglect patients. J Cogn Neurosci. *14*, 62-69.

Freeman,E. and Driver,J. (2008). Direction of visual apparent motion driven solely by timing of a static sound. Curr.Biol. *18*, 1262-1266.

Friedman,D., Cycowicz,Y.M., and Gaeta,H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neuroscience and Biobehavioral Reviews *25*, 355-373.

Friston,K.J., Penny,W., Phillips,C., Kiebel,S., Hinton,G., and Ashburner,J. (2002). Classical and Bayesian inference in neuroimaging: theory. Neuroimage *16*, 465-483.

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

Fu,K.M., Johnston,T.A., Shah,A.S., Arnold,L., Smiley,J., Hackett,T.A., Garraghty,P.E., and Schroeder,C.E. (2003). Auditory cortical neurons respond to somatosensory stimulation. J Neurosci. *23*, 7510-7515.

Fu,K.M., Shah,A.S., O'Connell,M.N., McGinnis,T., Eckholdt,H., Lakatos,P., Smiley,J., and Schroeder,C.E. (2004). Timing and laminar profile of eye-position effects on auditory responses in primate auditory cortex. J Neurophysiol. *92*, 3522-3531.

Gebhard,J.W., ., and Mowbray,G.H. (1959). On discriminating the rate of visual flicker and auditory flutter. Am.J.Psychol. *72*, 521-529.

Genovese,C.R., Lazar,N.A., and Nichols,T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. Neuroimage *15*, 870-878.

Ghazanfar,A.A., Maier,J.X., Hoffman,K.L., and Logothetis,N.K. (2005). Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. J Neurosci. *25*, 5004-5012.

Ghazanfar,A.A. and Schroeder,C.E. (2006). Is neocortex essentially multisensory? Trends Cogn Sci. *10*, 278-285.

Giard,M.H. and Peronnet,F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J Cogn Neurosci. *11*, 473-490.

Giard,M.H., Perrin,F., Pernier,J., and Bouchet,P. (1990). Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. Psychophysiology *27*, 627-640.

Gondan,M., Niederhaus,B., Rosler,F., and Roder,B. (2005). Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. Percept.Psychophys. *67*, 713-726.

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

Hodsoll,J., Mevorach,C., and Humphreys,G.W. (2009). Driven to less distraction: rTMS of the right parietal cortex reduces attentional capture in visual search. Cereb.Cortex *19*, 106-114.

Hotting,K., Rosler,F., and Roder,B. (2003). Crossmodal and intermodal attention modulate event-related brain potentials to tactile and auditory stimuli. Exp.Brain Res. *148*, 26-37.

Howard,I.P. and Templeton,W.B. (1966). Human Spatial Orientation. (London: John Wiley and Sons Ltd).

Huk,A.C. and Heeger,D.J. (2002). Pattern-motion responses in human visual cortex. Nat.Neurosci. *5*, 72-75.

Hyder,F., Rothman,D.L., Mason,G.F., Rangarajan,A., Behar,K.L., and Shulman,R.G. (1997). Oxidative glucose metabolism in rat brain during single forepaw stimulation: a spatially localized 1H[13C] nuclear magnetic resonance study. J.Cereb.Blood Flow Metab *17*, 1040-1047.

Jaaskelainen,I.P., Ahveninen,J., Bonmassar,G., Dale,A.M., Ilmoniemi,R.J., Levanen,S., Lin,F.H., May,P., Melcher,J., Stufflebeam,S., Tiitinen,H., and Belliveau,J.W. (2004). Human posterior auditory cortex gates novel sounds to consciousness. Proc.Natl.Acad.Sci.U.S.A *101*, 6809-6814.

Jonides,J. and Irwin,D.E. (1981). Capturing attention. Cognition *10*, 145-150.

Josephs O, Deichmann R, and Turner R (2000). Trajectory measurement and generalised reconstruction in rectilinear EPI. Proceedings of ISMRM *8*, 1517.

Jueptner,M. and Weiller,C. (1995). Review: does measurement of regional cerebral blood flow reflect synaptic activity? Implications for PET and fMRI. Neuroimage *2*, 148-156.

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

Kincade,J.M., Abrams,R.A., Astafiev,S.V., Shulman,G.L., and Corbetta,M. (2005). An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. J Neurosci. *25*, 4593-4604.

Kitagawa,N. and Ichihara,S. (2002). Hearing visual motion in depth. Nature *416*, 172-174.

Korzyukov,O.A., Winkler,I., Gumenyuk,V.I., and Alho,K. (2003). Processing abstract auditory features in the human auditory cortex. Neuroimage *20*, 2245-2258.

Kubovy,M. (1981). In Perceptual organization, M.Kubovy & J.R.Pomerantz, ed. Hillsdale, NJ: Erlbaum), pp. 55-98.

Kwong,K.K., Belliveau,J.W., Chesler,D.A., Goldberg,I.E., Weisskoff,R.M., Poncelet,B.P., Kennedy,D.N., Hoppel,B.E., Cohen,M.S., Turner,R., and . (1992).

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

Lakatos,P., Chen,C.M., O'Connell,M.N., Mills,A., and Schroeder,C.E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron *53*, 279-292.

Lange,K. and Roder,B. (2006). Orienting attention to points in time improves stimulus processing both within and across modalities. J.Cogn Neurosci. *18*, 715-729.

Lavie,N. and de Fockert,J. (2005). The role of working memory in attentional capture. Psychon.Bull.Rev *12*, 669-674.

Lemus,L., Hernandez,A., Luna,R., Zainos,A., and Romo,R. (2010). Do sensory cortices process more than one sensory modality during perceptual judgments? Neuron *67*, 335-348.

Liebenthal,E., Ellingson,M.L., Spanaki,M.V., Prieto,T.E., Ropella,K.M., and Binder,J.R. (2003). Simultaneous ERP and fMRI of the auditory cortex in a passive oddball paradigm. Neuroimage *19*, 1395-1404.

Lipschutz,B., Kolinsky,R., Damhaut,P., Wikler,D., and Goldman,S. (2002). Attention-dependent changes of activation and connectivity in dichotic listening. Neuroimage *17*, 643-656.

Lloyd,D.M., Shore,D.I., Spence,C., and Calvert,G.A. (2003). Multisensory representation of limb position in human premotor cortex. Nat.Neurosci. *6*, 17-18.

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

Lu,Z.L. and Dosher,B.A. (1998). External noise distinguishes attention mechanisms. Vision Res. *38*, 1183-1198.

Luck,S.J. and Hillyard,S.A. (1994). Spatial filtering during visual search: evidence from human electrophysiology. J.Exp.Psychol.Hum.Percept.Perform. *20*, 1000-1014.

Luck,S.J., Hillyard,S.A., Mouloua,M., and Hawkins,H.L. (1996). Mechanisms of visual-spatial attention: resource allocation or uncertainty reduction? J.Exp.Psychol.Hum.Percept Perform. *22*, 725-737.

Macaluso,E., Frith,C.D., and Driver,J. (2000). Modulation of human visual cortex by crossmodal spatial attention. Science *289*, 1206-1208.

Magistretti,P.J. and Pellerin,L. (1999). Cellular mechanisms of brain energy metabolism and their relevance to functional brain imaging. Philos.Trans.R.Soc.Lond B Biol.Sci. *354*, 1155-1163.

Martinez,A., Anllo-Vento,L., Sereno,M.I., Frank,L.R., Buxton,R.B., Dubowitz,D.J., Wong,E.C., Hinrichs,H., Heinze,H.J., and Hillyard,S.A. (1999). Involvement of striate and extrastriate visual cortical areas in spatial attention. Nat.Neurosci. *2*, 364-369.

McGurk,H. and MacDonald,J. (1976). Hearing lips and seeing voices. Nature *264*, 746-748.

Meredith,M.A., Nemitz,J.W., and Stein,B.E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. J Neurosci. *7*, 3215-3229.

Meredith,M.A. and Stein,B.E. (1983). Interactions among converging sensory inputs in the superior colliculus. Science *221*, 389-391.

Mesulam,M.M. (1998). From sensation to cognition. Brain *121 ( Pt 6)*, 1013-1052.

Miller,J. (1982). Divided attention: evidence for coactivation with redundant signals. Cognit.Psychol. *14*, 247-279.

Molholm,S., Martinez,A., Ritter,W., Javitt,D.C., and Foxe,J.J. (2005). The neural circuitry of pre-attentive auditory change-detection: an fMRI study of pitch and duration mismatch negativity generators. Cereb.Cortex *15*, 545-551.

Molholm,S., Ritter,W., Javitt,D.C., and Foxe,J.J. (2004). Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cereb.Cortex *14*, 452-465.

Molholm,S., Ritter,W., Murray,M.M., Javitt,D.C., Schroeder,C.E., and Foxe,J.J. (2002). Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res.Cogn Brain Res. *14*, 115-128.

Morein-Zamir,S., Soto-Faraco,S., and Kingstone,A. (2003). Auditory capture of vision: examining temporal ventriloquism. Brain Res.Cogn Brain Res. *17*, 154-163.

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

Mottonen,R., Krause,C.M., Tiippana,K., and Sams,M. (2002). Processing of changes in visual speech in the human auditory cortex. Brain Res.Cogn Brain Res. *13*, 417-425.

Muckli,L., Kohler,A., Kriegeskorte,N., and Singer,W. (2005). Primary visual cortex activity along the apparent-motion trace reflects illusory perception. PLoS.Biol. *3*, e265.

Murray,M.M., Michel,C.M., Grave,d.P., Ortigue,S., Brunet,D., Gonzalez,A.S., and Schnider,A. (2004). Rapid discrimination of visual and multisensory memories revealed by electrical neuroimaging. Neuroimage *21*, 125-135.

Murray,M.M., Molholm,S., Michel,C.M., Heslenfeld,D.J., Ritter,W., Javitt,D.C., Schroeder,C.E., and Foxe,J.J. (2005). Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb.Cortex *15*, 963-974.

Naatanen,R., Paavilainen,P., Tiitinen,H., Jiang,D., and Alho,K. (1993). Attention and mismatch negativity. Psychophysiology *30*, 436-450.

Neisser,U. and Becklen,R. (1975). Selective looking: Attending to visually specified events. Cognitive Psychology *7*, 480-494.

O'Connor,D.H., Fukui,M.M., Pinsk,M.A., and Kastner,S. (2002). Attention modulates responses in the human lateral geniculate nucleus. Nat.Neurosci. *5*, 1203-1209.

O'Craven,K.M., Downing,P.E., and Kanwisher,N. (1999). fMRI evidence for objects as the units of attentional selection. Nature *401*, 584-587.

Ogawa,S., Lee,T.M., Kay,A.R., and Tank,D.W. (1990a). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc.Natl.Acad.Sci.U.S.A *87*, 9868-9872.

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

Olson,I.R., Gatenby,J.C., and Gore,J.C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Brain Res.Cogn Brain Res. *14*, 129-138.

Opitz,B., Mecklinger,A., Friederici,A.D., and von Cramon,D.Y. (1999a). The functional neuroanatomy of novelty processing: Integrating ERP and fMRI results. Cerebral Cortex *9*, 379-391.

Opitz,B., Mecklinger,A., von Cramon,D.Y., and Kruggel,F. (1999b). Combining electrophysiological and hemodynamic measures of the auditory oddball. Psychophysiology *36*, 142-147.

Opitz,B., Rinne,T., Mecklinger,A., von Cramon,D.Y., and Schroger,E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. Neuroimage *15*, 167-174.

Paavilainen,P., Alho,K., Reinikainen,K., Sams,M., and Naatanen,R. (1991). Right hemisphere dominance of different mismatch negativities. Electroencephalogr.Clin.Neurophysiol. *78*, 466-479.

Padberg,J., Seltzer,B., and Cusick,C.G. (2003). Architectonics and cortical connections of the upper bank of the superior temporal sulcus in the rhesus monkey: an analysis in the tangential plane. J.Comp Neurol. *467*, 418-434.

Parton,A., Malhotra,P., and Husain,M. (2004). Hemispatial neglect. J.Neurol.Neurosurg.Psychiatry *75*, 13-21.

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

Pekkola,J., Ojanen,V., Autti,T., Jaaskelainen,I.P., Mottonen,R., Tarkiainen,A., and Sams,M. (2005). Primary auditory cortex activation by visual speech: an fMRI study at 3 T. Neuroreport *16*, 125-128.

Posner,M.I. and Cohen,Y. (1984). Components of visual orienting. In Attention and Performance X, H. Bouma and D. G. Bouwhuis, eds. pp. 55-66.

Posner,M.I. (1980). Orienting of attention. Q.J.Exp.Psychol. *32*, 3-25.

Pugh,K.R., offywitz,B.A., Shaywitz,S.E., Fulbright,R.K., Byrd,D., Skudlarski,P., Shankweiler,D.P., Katz,L., Constable,R.T., Fletcher,J., Lacadie,C., Marchione,K., and Gore,J.C. (1996). Auditory selective attention: an fMRI investigation. Neuroimage *4*, 159-173.

Raij,T., Uutela,K., and Hari,R. (2000). Audiovisual integration of letters in the human brain. Neuron *28*, 617-625.

Reisberg,D. (1978). Looking where you listen: visual cues and auditory attention. Acta Psychol.(Amst) *42*, 331-341.

Ress,D. and Heeger,D.J. (2003). Neuronal correlates of perception in early visual cortex. Nat.Neurosci. *6*, 414-420.

Ridderinkhof,K.R., Ullsperger,M., Crone,E.A., and Nieuwenhuis,S. (2004). The role of the medial frontal cortex in cognitive control. Science *306*, 443-447.

Rinne,T., Sarkka,A., Degerman,A., Schroger,E., and Alho,K. (2006). Two separate mechanisms underlie auditory change detection and involuntary control of attention. Brain Res. *1077*, 135-143.

Rockland,K.S. and Ojima,H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. Int.J Psychophysiol. *50*, 19-26.

Rockland,K.S. and Van Hoesen,G.W. (1994). Direct temporal-occipital feedback connections to striate cortex (V1) in the macaque monkey. Cereb.Cortex *4*, 300-313.

Roeber,U., Widmann,A., and Schroger,E. (2003). Auditory distraction by duration and location deviants: a behavioral and event-related potential study. Brain Res.Cogn Brain Res. *17*, 347-357.

Rushworth,M.F. and Taylor,P.C. (2006). TMS in the parietal cortex: updating representations for attention and action. Neuropsychologia *44*, 2700-2716.

Saito,D.N., Yoshimura,K., Kochiyama,T., Okada,T., Honda,M., and Sadato,N. (2005). Cross-modal Binding and Activated Attentional Networks during Audio-visual Speech Integration: a Functional MRI Study. Cereb.Cortex.

Sams,M., Kaukoranta,E., Hamalainen,M., and Naatanen,R. (1991). Cortical activity elicited by changes in auditory stimuli: different sources for the magnetic N100m and mismatch responses. Psychophysiology *28*, 21-29.

Schairer,K.S., Gould,H.J., and Pousson,M.A. (2001). Source generators of mismatch negativity to multiple deviant stimulus types. Brain Topogr. *14*, 117-130.

Schmahmann,J.D. and Pandya,D.N. (1991). Projections to the basis pontis from the superior temporal sulcus and superior temporal region in the rhesus monkey. J.Comp Neurol. *308*, 224-248.

Schneider,K.A. and Kastner,S. (2009). Effects of sustained spatial attention in the human lateral geniculate nucleus and superior colliculus. J.Neurosci. *29*, 1784-1795.

Schroeder,C.E. and Foxe,J. (2005). Multisensory contributions to low-level, 'unisensory' processing. Curr.Opin.Neurobiol. *15*, 454-458.

Schroeder,C.E. and Foxe,J.J. (2002). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Brain Res.Cogn Brain Res. *14*, 187-198.

Schroeder,C.E., Lindsley,R.W., Specht,C., Marcovici,A., Smiley,J.F., and Javitt,D.C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. J Neurophysiol. *85*, 1322-1327.

Schroger,E. (1994). Automatic detection of frequency change is invariant over a large intensity range. Neuroreport *5*, 825-828.

Schroger,E., Giard,M.H., and Wolff,C. (2000). Auditory distraction: event-related potential and behavioral indices. Clin.Neurophysiol. *111*, 1450-1460.

Schroger,E. and Widmann,A. (1998a). Speeded responses to audiovisual signal changes result from bimodal integration. Psychophysiology *35*, 755-759.

Schroger,E. and Widmann,A. (1998b). Speeded responses to audiovisual signal changes result from bimodal integration. Psychophysiology *35*, 755-759.

Schroger,E. and Widmann,A. (1998c). Speeded responses to audiovisual signal changes result from bimodal integration. Psychophysiology *35*, 755-759.

Schroger,E. and Wolff,C. (1998). Behavioral and electrophysiological effects of task-irrelevant sound change: a new distraction paradigm. Brain Res.Cogn Brain Res. *7*, 71-87.

Seiffert,A.E., Somers,D.C., Dale,A.M., and Tootell,R.B. (2003). Functional MRI studies of human visual motion perception: texture, luminance, attention and after-effects. Cereb.Cortex *13*, 340-349.

Sekiyama,K., Kanno,I., Miura,S., and Sugita,Y. (2003). Auditory-visual speech perception examined by fMRI and PET. Neurosci.Res. *47*, 277-287.

Sekuler,R., Sekuler,A.B., and Lau,R. (1997). Sound alters visual motion perception. Nature *385*, 308.

Serences,J.T., Shomstein,S., Leber,A.B., Golay,X., Egeth,H.E., and Yantis,S. (2005). Coordination of voluntary and stimulus-driven attentional control in human cortex. Psychol.Sci. *16*, 114-122.

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

Shams,L., Kamitani,Y., and Shimojo,S. (2000). Illusions. What you see is what you hear. Nature *408*, 788.

Shams,L., Kamitani,Y., and Shimojo,S. (2002). Visual illusion induced by sound. Brain Res.Cogn Brain Res. *14*, 147-152.

Shams,L., Kamitani,Y., Thompson,S., and Shimojo,S. (2001). Sound alters visual evoked potentials in humans. Neuroreport *12*, 3849-3852.

Shams,L., Ma,W.J., and Beierholm,U. (2005). Sound-induced flash illusion as an optimal percept. Neuroreport *16*, 1923-1927.

Shipley,T. (1964). AUDITORY FLUTTER-DRIVING OF VISUAL FLICKER. Science *145*, 1328-1330.

Shomstein,S. and Yantis,S. (2004). Control of attention shifts between vision and audition in human cortex. J Neurosci. *24*, 10702-10706.

Shomstein,S. and Yantis,S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. J.Neurosci. *26*, 435-439.

Sperling,G., Budiansky,J., Spivak,J.G., and Johnson,M.C. (1971). Extremely rapid visual search: the maximum rate of scanning letters for the presence of a numeral. Science *174*, 307-311.

Stein,B.E. (1978). Development and organization of multimodal representation in cat superior colliculus. Fed.Proc. *37*, 2240-2245.

Stein,B.E. and Arigbede,M.O. (1972). Unimodal and multimodal response properties of neurons in the cat's superior colliculus. Exp.Neurol. *36*, 179-196.

Stein,B.E., London,N., Wilkinson,L., and Price,D. (1996). Enhancement of perceived visual intensity by auditory stimuli: a psychophysical analysis. J Cogn Neurosci. *8*, 497-506.

Stein,B.E., Magalhaes-Castro,B., and Kruger,L. (1975). Superior colliculus: visuotopic-somatotopic overlap. Science *189*, 224-226.

Stein,B.E., Stanford,T.R., Wallace,M.T., Vaughan,J.W., and Jiang,W. (2004). Crossmodal spatial interactions in subcortical and cortical circuits. In Crossmodal Space and Crossmodal Attnetion, C. Spence and J. Driver, eds. Oxford University Press), pp. 25-50.

Sterzer,P., Haynes,J.D., and Rees,G. (2006). Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT+/V5. Neuroimage. *32*, 1308-1316.

Sumby,WE. and Pollack,I. (1954). Visual contribution to speech intelligibility in noise. Journal of the Acoustical Soceity of America.

Talairach Tournoux (1988). Co-Planar Sterotaxic Atlas of the Human Brain. (Theime, Stuggart, Germany.

Talsma,D., Senkowski,D., Soto-Faraco,S., and Woldorff,M.G. (2010). The multifaceted interplay between attention and multisensory integration. Trends Cogn Sci. *14*, 400-410.

Teo,P.C., Sapiro,G., and Wandell,B.A. (1997). Creating connected representations of cortical gray matter for functional MRI visualization. IEEE Trans.Med.Imaging *16*, 852-863.

Theeuwes,J. (1991). Exogenous and endogenous control of attention: the effect of visual onsets and offsets. Percept.Psychophys. *49*, 83-90.

Theeuwes,J. (1992). Perceptual selectivity for color and form. Percept.Psychophys. *51*, 599-606.

Theeuwes,J. (1994). Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets. J Exp.Psychol.Hum.Percept.Perform. *20*, 799-806.

Theeuwes,J. (2010). Top-down and bottom-up control of visual selection. Acta Psychol.(Amst).

Theeuwes,J. and Burger,R. (1998). Attentional control during visual search: the effect of irrelevant singletons. J.Exp.Psychol.Hum.Percept.Perform. *24*, 1342-1353.

Tong,F. (2003). Primary visual cortex and visual awareness. Nat.Rev.Neurosci. *4*, 219-229.

Tootell,R.B., Reppas,J.B., Kwong,K.K., Malach,R., Born,R.T., Brady,T.J., Rosen,B.R., and Belliveau,J.W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J.Neurosci. *15*, 3215-3230.

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

Tzourio,N., Massioui,F.E., Crivello,F., Joliot,M., Renault,B., and Mazoyer,B. (1997). Functional anatomy of human auditory attention studied with PET. Neuroimage *5*, 63-77.

Van der Burg,E., Olivers,C.N., Bronkhorst,A.W., and Theeuwes,J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. J.Exp.Psychol.Hum.Percept.Perform. *34*, 1053-1065.

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

Villringer,A. and Dirnagl,U. (1995). Coupling of brain activity and cerebral blood flow: basis of functional neuroimaging. Cerebrovasc.Brain Metab Rev. *7*, 240-276.

von Grunau,M.W. (1986). A motion aftereffect for long-range stroboscopic apparent motion. Percept.Psychophys. *40*, 31-38.

von Kriegstein,K., Kleinschmidt,A., Sterzer,P., and Giraud,A.L. (2005). Interaction of face and voice areas during speaker recognition. J Cogn Neurosci. *17*, 367-376.

Vroomen,J. and de Gelder,B. (2004). Temporal ventriloquism: sound modulates the flash-lag effect. J.Exp.Psychol.Hum.Percept.Perform. *30*, 513-518.

Vroomen,J., de Gelder,B., and Vroomen,J. (2004). Temporal ventriloquism: sound modulates the flash-lag effect. J.Exp.Psychol.Hum.Percept.Perform. *30*, 513-518.

Wallace,M.T., Meredith,M.A., and Stein,B.E. (1998). Multisensory integration in the superior colliculus of the alert cat. J Neurophysiol. *80*, 1006-1010.

Wallace,M.T., Perrault,T.J., Jr., Hairston,W.D., and Stein,B.E. (2004). Visual experience is necessary for the development of multisensory integration. J.Neurosci. *24*, 9580-9584.

Wallace,M.T. and Stein,B.E. (1996). Sensory organization of the superior colliculus in cat and monkey. Prog.Brain Res. *112*, 301-311.

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

Watkins S, Shams,L., Josephs O, and Rees,G. (2007). Activity in human V1 follows multisensory perception. Neuroimage *17*, 572-578.

Watkins,S., Dalton,P., Lavie,N., and Rees,G. (2007). Brain mechanisms mediating auditory attentional capture in humans. Cereb.Cortex *17*, 1694-1700.

Watkins,S., Shams,L., Tanaka,S., Haynes,J.D., and Rees,G. (2006). Sound alters activity in human V1 in association with illusory visual perception. Neuroimage. *31*, 1247-1256.

Weiskopf N, Sitaram R, Josephs O, Veit R, Scharnowski F, Goebel R, Birbaumer N, Deichmann R, and Mathiak K (2006). Real-time functional magnetic resonance imaging: methods and applications. Magnetic Resonance Imaging *in press*.

Wolff,C. and Schroger,E. (2001). Activation of the auditory pre-attentive change detection system by tone repetitions with fast stimulation rate. Brain Res.Cogn Brain Res. *10*, 323-327.

Yantis,S. (1993). Stimulus-driven attentional capture and attentional control settings. J Exp.Psychol.Hum.Percept.Perform. *19*, 676-681.

Yantis,S. and Egeth,H.E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. J Exp.Psychol.Hum.Percept.Perform. *25*, 661-676.

Zatorre,R.J., Mondor,T.A., and Evans,A.C. (1999). Auditory attention to space and frequency activates similar cerebral systems. Neuroimage *10*, 544-554.

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