THE ANATOMICAL SUBSTRATES OF FEATURE INTEGRATION DURING OBJECT PROCESSING

Julia Hocking

Wellcome Trust Centre for Neuroimaging
Institute of Neurology
University College London

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

For Cathy, with love

ABSTRACT OF THESIS

Objects can be identified from a number of perceptual attributes, including visual, auditory and tactile sensory input. The integration of these perceptual attributes constitutes our semantic knowledge of an object representation. This research uses functional neuroimaging to investigate the brain areas that integrate perceptual features into an object representation, and how these regions are modulated by stimulus- and task-specific features.

A series of experiments are reported that utilise different types of perceptual integration, both within and across sensory modalities. These include 1) the integration of visual form with colour, 2) the integration of visual and auditory object features, and 3) the integration of visual and tactile abstract shapes. Across these experiments I have also manipulated additional factors, including the meaning of the perceptual information (meaningful objects versus meaningless shapes), the verbal or non-verbal nature of the perceptual inputs (e.g. spoken words versus environmental sounds) and the congruency of crossmodal inputs.

These experiments have identified a network of brain regions both common to, and selective for, different types of object feature integration. For instance, I have identified a common bilateral network involved in the integration and association of crossmodal audiovisual objects and intra-modal auditory or visual object pairs. However, I have also determined that activation in response to the same concepts can be modulated by the type of stimulus input (verbal versus nonverbal), the timing of those inputs (simultaneous versus sequential presentation), and the congruency of stimulus pairs (congruent versus incongruent). Taken together, the results from these experiments demonstrate modulations of neuronal activation by different object attributes at multiple different levels of the object processing hierarchy, from early sensory processing through to stored object representations. Critically, these differential effects have even been observed with the same conceptual stimuli. Together these findings highlight the need for a model of object feature processing that can account for the functional demands that elicit these anatomical differences.

TABLE OF CONTENTS

Abstract **3**

Table of Contents **4**

List of Figures **8**

List of Tables **9**

Acknowledgements **10**

1 General introduction **13**

1.1 Introduction *13*

1.2 Multisensory integration in cortical and subcortical structures *15*

1.3 Integration of audiovisual linguistic/phonetic information *17*

1.4 Integration of audiovisual conceptual/semantic information *23*

1.5 Integration of tactile-visual information *28*

1.6 Overview of following chapters *33*

2 Methods for acquisition and analysis of functional imaging data **38**

2.1 Introduction *38*

2.2 Positron Emission Tomography *38*
2.21 Principles of PET 38
2.22 Data Acquisition 40

2.3 Functional Magnetic Resonance Imaging *41*
2.31 Principles of fMRI 41

2.4 Data Analysis using Statistical Parametric Mapping *43*
2.41 Pre-processing 44
2.42 Realignment 45
2.43 Normalisation 46
2.44 Smoothing 47

2.5 The General Linear Model *47*
2.51 Contrasts and Inference 48

3 Examining the effect of increased perceptual cues on object naming **52**

3.1 Introduction *52*

3.2 Method *55*
3.21 Materials and Methods 55
3.22 Behavioural Study 58
3.23 Data Acquisition 59

3.3 Behavioural Results *60*

3.4 Functional Imaging Results 61
3.41 Whole brain analysis 61
3.42 Regions of interest 62

3.5 Discussion 65
3.51 Visual perceptual processing 66
3.52 Semantic processing 67
3.53 Crossmodal audiovisual processing. 69

3.6 Summary 70

4 The role of the posterior superior temporal sulcus in auditory and visual object processing 72

4.1 Introduction 72
4.11 Bimodal versus uni-modal stimuli 78
4.12 Effect of task 79
4.13 Congruent versus incongruent matching 80

4.2 Method 83
4.21 Materials and Method 83
4.22 Data Acquisition 87

4.3 Behavioural Results 90

4.4 Functional imaging analysis 91
4.41 Intra-modal matching 91
4.42 Crossmodal versus Intra-modal matching 92
4.43 Effect of Congruency 92
4.44 Summary 93

4.5 Discussion 95
4.51 Bimodal versus Uni-modal object matching 95
4.52 Effect of Task 96
4.53 Congruent versus Incongruent trials 97
4.54 Patchy organisation within human STS 97

4.6 Summary 98

5 The influence of congruency and presentation rate on audiovisual object processing 101

5.1 Introduction 101

5.2 Method 104
5.21 Materials and Methods 104

5.3 Behavioural Results 107

5.4 Functional Imaging Results 108
5.41 The effect of congruency during simultaneous and sequential crossmodal matching 108
5.42 The effect of simultaneous versus sequential crossmodal matching 110

5.5 Discussion 111
5.51 Congruent versus incongruent matching 112

5.52 Simultaneous versus Sequential Presentation 113

5.6 Summary 114

6 Verbal versus nonverbal audiovisual object processing 116

6.1 Introduction 116
6.11 Dissociating verbal and nonverbal processing 118
6.12 This study 120

6.2 Method 121
6.21 Materials and Method 121
6.22 Data Acquisition 122

6.3 Behavioural Results 123

6.4 Functional Imaging Results 124
6.41 The effect of verbal information on audiovisual matching 125
6.42 The effect of nonverbal input on audiovisual matching 127

6.5 Discussion 129
6.51 Verbal audiovisual matching 129
6.52 Nonverbal audiovisual matching 130

6.6 Summary 133

7 Crossmodal integration of tactile-visual abstract shapes 135

7.1 Introduction 135
7.11 Findings from non-human primate studies 136
7.12 Neuropsychological data in humans 136
7.13 Functional imaging of normal subjects 137
7.14 This Study 138

7.2 Methods 139
7.21 Materials and Methods 139
7.22 Data Acquisition 142
Data Analysis 143

7.3 Behavioural Results 144

7.4 Functional Imaging analysis 145
7.41 Intra-modal matching 145
7.42 Crossmodal versus intra-modal matching 145
7.43 Congruent versus incongruent matching 145

7.5 Discussion 146
7.51 Crossmodal versus intra-modal matching in the insula 147
7.52 Integration in intraparietal sulcus 148
7.53 Integration in LOtv 150
7.54 Integration in medial temporal lobes 151

7.6 Summary 152

8 Crossmodal object processing in the lateral and medial anterior temporal cortex 155

8.1 Introduction 155

8.2 Method 158
  8.21 Materials and Methods 158
  8.22 Data Analysis 160

8.3 Behavioural results 161

8.4 Functional Imaging Results 161
  8.41 Congruent > Incongruent Trials (over experiments) 161
  8.42 Incongruent > Congruent Trials (over experiments) 162
  8.43 Interaction between congruency and experiment 162
  8.44 Summary of key findings in left anterior temporal cortex 165

8.5 Discussion 166
  8.51 Increased activation for congruent stimuli in the left temporal pole 166
  8.52 Activation for incongruent objects in left infero-lateral temporal cortex 168
  8.53 Activation by crossmodal inputs in antero-medial temporal lobe 170

8.6 Summary 173

9 Conclusions 176

9.1 Summary of thesis 176

9.2 Experimental findings 176

9.3 Crossmodal processing at multiple levels 181
  9.31 Posterior superior temporal sulcus 183
  9.32 Anterior temporal lobe 184
  9.33 Multiple regions, processes and levels 188

9.4 Experimental factors 189
  9.41 Level of stimulus input 189
  9.42 Congruency 190
  9.43 Synchrony 191

9.5 Final Conclusions 191

10 References 195

11 Appendices 229
  Appendix 1 229
  Appendix 2 230
  Appendix 3 231
  Appendix 4 234

I, Julia Hocking, 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.

LIST OF FIGURES

Figure 2.1: *Signal detection in a PET camera* 39

Figure 2.2: *Function for modelling the BOLD component of the haemodynamic response* 43

Figure 2.3: *Stages of image processing and statistical analysis for PET and fMRI data, as implemented in SPM2* 44

Figure 3.1: *Example of a single stimulus trial for each of the six conditions* 58

Figure 3.2: *Main effect of increased perceptual input* 64

Figure 3.3: *Interaction in antero-medial temporal lobe* 64

Figure 4.1: *Anatomical location of the region of interest* 73

Figure 4.2: *Stimulus trials for Experiments 1a and 1b* 87

Figure 4.3: *Main effect of intra-modal matching* 91

Figure 4.4: *A common system for matching crossmodal and intra-modal stimuli* 94

Figure 5.1: *Stimulus trials for simultaneous presentation in Experiment 2* 106

Figure 5.2: *Incongruent > congruent matching of simultaneously presented audiovisual objects* 109

Figure 5.3: *Effect sizes in bilateral pSTS for all conditions in each Experiment* 110

Figure 5.4: *Comparison of audiovisual matching in Experiments 1 and 2* 111

Figure 6.1: *Stimulus trials for simultaneous audiovisual matching* 122

Figure 6.2: *Common network for audiovisual integration* 125

Figure 6.3: *Effect of verbal input on audiovisual matching* 126

Figure 6.4: *Effect of no verbal input on audiovisual matching* 128

Figure 7.1: *Sample stimuli used during each matching condition* 141

Figure 7.2: *Main effect of intra-modal matching* 145

Figure 7.3: *Crossmodal versus intra-modal matching* 146

Figure 8.1: *Summary of key findings in left anterior temporal cortex*
a: *Congruent > incongruent trials in left temporal pole* 164
b: *Incongruent > congruent in left lateral inferior temporal* 164
c: *Interaction of experiment with congruency* 165

LIST OF TABLES

Table 3.1: *Naming latencies (mean and standard deviation) for behavioural experiment* 61

Table 3.2: *Regional activation for main effect of increased perceptual input* 63

Table 4.1: *Anatomical co-ordinates of temporal regions*
a: Audiovisual effects in posterior STS 74
b: Audiovisual effects in anterior STS 74
c: Semantic and sentence effects in temporal regions 75

Table 4.2: *Behavioural data* 91

Table 5.1: *Behavioural data* 108

Table 5.2: *Incongruent relative to congruent trials for simultaneous matching* 109

Table 6.1: *Behavioural data* 124

Table 6.2: *Main effect of maximum verbal information* 126

Table 6.3: *Simple main effects for verbal information in posterior STS* 127

Table 6.4: *Main effect of nonverbal matching* 128

Table 6.5: *Nonverbal effects in the right fusiform* 129

Table 7.1: *Response accuracy for tactile-visual matching* 144

Table 8.1: *Main effect of congruency* 162

Table 8.2: *Interaction between congruency and experiment* 163

ACKNOWLEDGEMENTS

First of all I would like to thank all those people at the Functional Imaging Laboratory that made the work in this thesis possible. Most importantly, Cathy for giving me the fantastic opportunity to study with her, first for the research component of my Masters degree, and for having the belief in me to continue her support throughout her supervision of my PhD. Thanks also go to the Wellcome Trust for funding my studentship.

The atmosphere created and nurtured at the FIL is unique. On the one hand, the scientific endeavour of all its researchers, led by the Principle Investigators, is at the cutting edge of functional brain imaging across the world. On the other, it is a place that allows individuals to grow both professionally and personally, within a culture of support and care. This unique balance provides an environment that at all times I felt privileged to be a part of.

Without order of preference, there are a number of people at the FIL I would like to mention. I would like to thank the functional technologies team – Amanda, Jan and David - for their good humour, patience and experience; Peter for always having a solution; Eric for his programming skills (and karaoke); and Ric for his quick response to IT problems (and for introducing me to The West Wing). I would like to thank my office colleagues for willingly providing answers to SPM (and many other) questions; for their company on much needed breaks to Caffe Nero; and for many enjoyable and stimulating conversations at The Queens Larder. Over the time that I was at the FIL, the size of the weekly language group meetings grew from Cathy, Andrea, Jenny and Uta and myself, to a much larger group. Regardless of the number of people, the atmosphere was always one where all voices were equal, questions were posed, research was shared, and problems were aired and (usually!) solved. These meetings were always a great way to start the week, and ensured you always felt part of a supportive group, thanks to the unique style that Cathy brings to her work. Monday mornings have not been the same since, and I miss all the lovely people, past and present, from the language group.

On a personal note, the long journey that has brought me to this point in my career started with a decision 10 years ago that couldn't have been carried out without the unequivocal support of my partner, Mike. Throughout my years of studying, the belief that you instilled in me that I could achieve my dreams has kept me on this enjoyable and rewarding path. Even though we are no longer together, I still hear your voice encouraging, supporting and believing in me. I can never thank you enough.

To Neri for your special friendship, love and support. I can't mention everything you've done for me (no really, I can't - all the red wine we've shared must have impaired my memory!), but I hope I can just say a very special thank you, to you (and Nick), from the bottom of my heart.

To Mum and Dad - thank you for your love, support and encouragement. Any success I have is a reflection of the foundations that you provided for me. I wish that Australia weren't so far away from you.

I found myself finishing this thesis on the other side of the world, away from all that was familiar and comfortable. My newfound friends have been wonderfully patient waiting for me to finish, and always enthusiastic with their praise of even the smallest achievements. Special thanks to Katie, Matt, Emma and of course Peter. Big love to my thesis-writing buddy Lucy (not long now, Dr Lulittle). You are my rock.

Having such a dedicated, supportive, understanding and loving supervisor has made this journey possible. Cathy, you are a very special person. Over the past years you've helped me through what has felt, at times, to be a rollercoaster of life events, for which I am eternally grateful. Your energy, enthusiasm, sense of fun, and naturally joyful spirit are some of the many positive qualities that make you such a fantastic person to be around. I'm proud of the work that we have achieved, and I'm proud that I can also call you my friend.

Chapter 1

1 General introduction

1.1 Introduction

Under normal circumstances, we recognise objects using not only our eyes, but using the convergent inputs from each relevant sense, or the sense that provides us with the most salient information. To be able to fully understand how our brain responds to, accesses and stores the available sensory information provided by our environment, empirical investigations must develop and expand current knowledge of object processing to include the integration of these different senses. This thesis uses positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) to investigate the integration of perceptual and conceptual object information from different input modalities. Modality is used here to refer to sensory input - in this thesis these inputs are in the auditory, visual and tactile modalities.

Clearly, in the traditional scanning environment, gaining the ecological validity of day to day object experience presents some insurmountable challenges. However, within the constraints of the technique, it is possible to present different modalities of perceptual input to participants both separately and simultaneously, and examine the way that these different inputs modulate neuronal activity. This can be especially enlightening to theoretical discussions of semantic organisation if the same object concepts elicit differential activation depending upon the sensory input or the combination of inputs, or if shared regions respond equivalently to inputs from different senses.

Multiple studies have shown that neuronal responses underlying access to stored object knowledge interact with many factors, including the demands of an experimental task or type of perceptual input. In the main, these observations have been based on paradigms that engage brain regions in response to one input modality. By presenting stimuli not only in one sensory input modality, but in different combinations of modalities, a richer functional-anatomical picture of object processing regions will be elucidated. In particular, the use of multi-modal stimuli can be used to dissociate regions which respond to combinations of different sensory inputs over and above their uni-modal counterparts (i.e. crossmodal or hetero-modal integration areas), and amodal areas where the response is independent of sensory input modality.

In Chapters 3, 4, 5 and 6 of this thesis I will present three studies designed to consider the functional levels at which auditory and visual object integration occurs. The goal of this group of studies is to investigate whether 1) integration of conceptual inputs occurs within an amodal semantic system, 2) is combined at an earlier sensory or perceptual level prior to accessing semantics, and 3) whether there are fundamental differences in how these crossmodal compared to uni-modal cues are integrated. These questions will be addressed through the manipulation of task (naming versus semantic decision), type of material (verbal versus nonverbal), sensory input modality (visual versus auditory) and presentation rate (simultaneous versus successive inputs).

The second part of this thesis (Chapters 7 and 8) examines the integration of abstract tactile and visual object stimuli. The first experiment asks whether crossmodal integration of abstract tactile-visual stimuli engages different regions compared to uni-modal tactile and visual stimuli. Following that, a combined analysis of different types of integration (audiovisual and tactile-visual), addresses the issue of whether there are brain regions commonly engaged by both types of integration. If so, then the functional processes that drive responses in these shared regions need to be considered.

In this introductory chapter I will review our current knowledge of how crossmodal objects are processed and represented in the brain, and how regional activation within a shared distributed network is modulated through the manipulation of a number of different experimental factors. In the sections that follow I will review evidence of four different kinds of crossmodal integration. This will serve to illustrate the different functional and anatomical levels at which crossmodal stimulus inputs can be combined. Particular attention is paid to the critical role that temporal lobe structures play in the processing and representation of amodal conceptual knowledge, within which integration of perceptual and semantic crossmodal inputs may occur. This section will also include discussion of the anatomical regions implicated in the functional integration of crossmodal inputs that are used as regions of interest revisited in many of the experimental chapters that follow. This introduction will then finish with a description and relevant aims of each experimental chapter.

1.2 Multisensory integration in cortical and subcortical structures

Studies of non-human primates have shown that neurons in many parts of the brain receive converging inputs from different uni-modal sensory systems.

Electrophysiological studies have demonstrated that cells in these convergence regions are responsive to stimulation from more than one modality. For example, Stein *et al.* (1975) identified overlapping maps in the subcortical superior colliculus for auditory, visual and tactile inputs that arose from the same (external) spatial location. When two or more of these sensory cues were presented in close temporal and spatial proximity, the response of these neurons was enhanced over and above the firing rate expected when summing responses to the two corresponding uni-modal inputs. In contrast, spatially disparate crossmodal inputs produce response depression in the same cells (Kadunce *et al.*, 1997). Neuroanatomical studies of non-human primates have also identified potential cortical regions where auditory, visual and tactile sensory afferents converge. These include the anterior and posterior superior temporal sulcus (a/pSTS: Bruce *et al.*, 1981; Desimone and Gross, 1979; Desimone and Ungerleider, 1986), ventral and lateral parietal cortex (Bremmer *et al.*, 2001; Lewis and Van Essen, 2000; Linden *et al.*, 1999) and the medial temporal lobe, including the rhinal cortices and amygdala (Friedman *et al.*, 1986; Murray and Mishkin, 1985; Suzuki, 1996; Suzuki and Amaral, 1994a, 1994b).

From these neuroanatomical and electrophysiological studies, principles governing synthesis of sensory cues for multisensory integration have been identified. These include sensitivity to temporal and spatial correspondence, response enhancement and depression, and the rule of inverse effectiveness, where responses to crossmodal inputs are maximal when responses to individual stimuli are minimally effective. Do these principles also apply to human studies? Evidence increasingly suggests that these defined response properties depend upon the type of sensory information to be integrated, as well as task-related cognitive factors. For instance, although temporal and spatial correspondence were believed to be critical for modulation of crossmodal effects based on non-human studies (e.g. Stein and Meredith, 1993), temporal correspondence is not critical for perception of the phonetically-mediated audiovisual McGurk effect (Munhall *et al.*, 1996) and audio-somatosensory interactions are not dependent upon spatial alignment (Murray *et al.*, 2005). Moreover, audiovisual and visual-somatosensory interactions between multisensory inputs have been reported not only in distributed shared neural systems, but also in the non-preferred uni-sensory cortices (Bushara *et al.*, 2003; Murray *et al.*, 2004). Attention has also been found to modulate processes involved in multisensory integration (Talsma *et al.*, 2006), where neuronal response enhancement was only observed when subjects attended to both auditory and visual inputs. It is therefore important to determine whether the cortical regions defined under the principles derived from non-human animals apply to perceptual stimuli such as lexical or semantically-mediated inputs in human subjects, and at what level this crossmodal interaction or integration takes place.

1.3 Integration of audiovisual linguistic/phonetic information

Crossmodal integration of linguistic information can be divided into those studies that have investigated the processing of spoken letters/phonemes and those that have used whole word and/or sentence processing, in the context of visual input presenting a moving mouth. Where the former can be considered the study of abstract, arbitrary sound-symbol associations, the latter stimuli can be ascribed meaning either in terms of the conceptual association derived from a concrete noun or from a syntactically governed series of words depicting an event or description. These studies have manipulated factors such as synchronisation (versus desynchronisation) of audiovisual inputs, congruency of the combined auditory and visual inputs, and attention to one or other of the crossmodal signals. Across virtually all of these studies, the focus of interest for the site of integration of these audiovisual inputs has been the posterior superior temporal sulcus (pSTS). In this section, the role of this region in the integration of phonetic and/or linguistic stimuli is reviewed, and reference is made to additional regions (the insula and anterior temporal cortex) that require further experimental exploration.

Calvert and colleagues (Calvert *et al.*, 1997, 1999, 2000, 2001) extended the principles of crossmodal integration derived from non-human studies to functional imaging of normal subjects using speech sounds and visual lip movements. Initially they observed an enhancement during audiovisual speech in auditory and visual uni-modal cortices, over and above the activation observed for uni-modal speech (Calvert *et al.*, 1999), but no additional multisensory regions. These findings are consistent with event-related potential (ERP) data showing modulation of secondary auditory and visual cortices by the non-preferred sensory modality (e.g. Giard and Peronnet, 1999; Molholm *et al.*, 2002; Murray *et al.*, 2004; Sams *et al.*, 1991). However, in a later study where subjects passively listened to audiovisual excerpts of prose (relative to uni-modal auditory and visual inputs), a super-additive enhancement effect for audiovisual speech was seen in pSTS when the audio and visual inputs were congruent (i.e. in synchrony) and relative suppression in pSTS for incongruent signals (Calvert *et al.*, 2000). These super- and sub-additive effects for multi-modal relative to uni-modal inputs are the defining criteria for an area involved in the active integration of multi-modal inputs, as opposed to an "amodal" area that responds independently of the sensory input. However, as I will argue in the next paragraph, super- and sub-additive activation for multi-modal relative to uni-modal inputs does not necessarily imply an active integration effect. These effects can also emerge from differences in intelligibility, attention and total sensory/perceptual input. Therefore, one of the main questions addressed in this thesis is whether pSTS is a site where effective integration of audiovisual signals occurs, or whether it is in fact a region responding to the final (meaningful) speech percept. In other words, this implies involvement of the pSTS in responding to linguistic/phonetic or semantic signals at a later stage of processing, rather than being a site of early multisensory integration (see for example Binder, 1997; Humphries *et al.*, 2006; Narain *et al.*, 2003).

The distinction between "multi-sensory integration" and "amodal processing" rests on determining whether activation is higher for multi-modal inputs than the sum of uni-modal inputs (i.e. a super-additive effect) when all other factors are controlled. In the Calvert *et al.* (2000) study reported above, attention, total sensory input and intelligibility were not controlled, which calls into question whether the super- and sub-additive effects they report in the pSTS are truly a reflection of multi-sensory integration. Briefly, when subjects are presented with incongruent speech, intelligibility is reduced because the visual inputs conflict with the auditory inputs. Consequently, activation in an amodal area sensitive to intelligibility of the speech percept will be lower for incongruent audiovisual speech relative to congruent audiovisual speech or uni-modal speech. This would result in a sub-additive, or negative interaction effect (incongruent $AV < [A + V]$) that does not reflect multi-sensory integration. Conversely, a super-additive effect ($AV > [A + V]$) could arise in an amodal processing area that was sensitive to the total sensory input at a given time point. This is because sensory inputs in audiovisual speech are twice those in uni-modal speech. Moreover, the super-additive effect may only be observed for congruent audiovisual speech because of attention and intelligibility confounds during incongruent speech. Specifically, to maximize intelligibility during incongruent audiovisual speech, subjects need to minimise attention to the conflicting visual inputs. Therefore perceptual processing on any given trial will be less for incongruent than congruent audiovisual speech. These issues are addressed in more detail in Chapter 4, where other experiments using congruency effects to identify integration regions are reviewed. To summarise the evidence so far, pSTS does not appear to be necessary for multisensory integration based on Calvert *et al.*, (1999) and is confounded by intelligibility, attentional, and perceptual processing factors in their later experiment (Calvert *et al.*, 2000).

Behaviourally, speech comprehension is improved when the speaker's moving mouth can be seen as well as heard (Rosen *et al.*, 1980), yet crossmodal interactions can alter as well as enhance perception of inputs. For instance, watching a person mouthing a syllable, and hearing a different syllable can lead to the McGurk effect (McGurk and MacDonald, 1976), where the combined signal leads to perception of a different syllable from that provided by the actual stimuli. Robust effects for the modulation of auditory cortex during the viewing of silent mouth movements (i.e. lip-reading: Callan *et al.*, 2001; Calvert *et al.*, 1997; MacSweeney *et al.*, 2000) have raised questions about the route through which visual and auditory speech cues are integrated. Two studies looking specifically at audiovisual linguistic integration have exploited the McGurk effect as a means by which to investigate discrepant visual input on auditory speech perception (Olson *et al.*, 2002; Sekiyama *et al.*, 2003), the assumption being that regions activated in response to the perception of audiovisual input relative to uni-modal speech or visual input are involved in the integration of the crossmodal signals.

By manipulating the intelligibility of the audiovisual signal through changes in the amplitude of the auditory signal (low versus high intelligibility based on decibel level), Sekiyama *et al.* (2003) observed increased activation in the pSTS for the low versus high intelligibility condition, but no difference between auditory and audiovisual conditions during high intelligibility. The pSTS has been implicated many times as a site for audiovisual integration (see Calvert, 2001; Calvert and Lewis, 2004 for reviews), and the role of this region will be returned to many times through the course of this thesis. The data presented by Sekiyama *et al.* (2003) raise the issue that pSTS is activated when subjects need to attend more strongly to the visual input, in order to supplement the low intelligibility of the auditory signal. In contrast, when intelligibility is high, subjects did not attend to the visual input, and behaviourally did not significantly experience the McGurk effect.

Rather than manipulating intelligibility, Olson *et al.* (2002) used the comparison of synchronised and desynchronised audiovisual speech to look at regions involved in integration using the McGurk effect relative to lip-reading words. The McGurk effect would only be expected in the synchronised condition (due to the onset asynchrony offset being set at 1 second for desynchronised inputs), an effect that was confirmed behaviourally post-scanning. Despite the behavioural differences in speech perception between synchronised and desynchronised speech, the regions activated were virtually identical, including the pSTS and posterior superior temporal gyrus (pSTG). The only regions differentially activated for synchronised > desynchronised audiovisual speech were observed in the left insula/claustrum and the left temporal pole. The insula has been reported in studies of both tactile-visual integration (Banati *et al.,* 2000; Hadjikhani and Roland, 1998: see below), and audiovisual integration (Bushara *et al.,* 2001; Calvert *et al.*, 2001), its role in multisensory integration suggested as a subcortical relay station due to its location (in non-human primates) receiving afferents, and projecting to, auditory, visual and somatosensory systems (Pearson *et al.,* 1982). Olson *et al.* (2002) do not discuss the additional activation observed in the left temporal pole, but the wider literature sees this region either as a crossmodal "hub" mediating communication amongst modality specific semantic representations (Rogers *et al.,* 2005, 2006), or alternatively a semantically-mediated amodal convergence region for unique items (Damasio, 1989; Tranel *et al.*, 1997). This region will be discussed further below.

The study by Olson *et al.* (2002) raises an interesting question with regard to the temporal alignment principle of multisensory integration. Temporal alignment has been advanced as a requirement for measuring multisensory integration in primates (Stein *et al.*, 1993; Kadunce *et al.,* 1997). However, Olson *et al.* (2002) found no difference in pSTS (and pSTG) for trials with or without corresponding (i.e. temporally aligned) mouth movements, thereby questioning the application of this principle to human studies.

In summary, the studies reviewed here, and those described in more detail in subsequent chapters of this thesis, suggest a modulation of activation depending upon the congruency of audiovisual inputs and the timing of those cues. The issues of congruency and presentation timing are considered in chapters 4 and 5 where task has been controlled across these different parameters. Moreover, although there has been a focus to date on the response properties of pSTS, there are clearly alternative regions that need to be functionally defined in terms of their role in crossmodal integration. These regions will be investigated in this thesis through the manipulation of different types of crossmodal integration. The insula has been reported in studies of tactile-visual integration (see section 1.5 below), supporting data that will be presented in Chapter 7. In Chapter 8, a direct comparison between crossmodal cues from perceptual (tactile-visual) and conceptual (audiovisual) inputs is made, and specifically investigates the modulating role that the anterior temporal cortices play in the processing of these different types of integration.

1.4 Integration of audiovisual conceptual/semantic information

The foundation that has been the primary driving force for the data contained in this thesis is the near absence of functional imaging studies investigating the integration of audiovisual information at a conceptual rather than early sensory or lexical/phonological level. Despite a large number of functional imaging studies looking at neural regions engaged when processing pictures and written word inputs or environmental sounds and spoken words as single inputs, only a small handful of studies have combined these types of information within a single experiment. There are two studies particularly relevant to a discussion specifically on crossmodal integration of conceptual information, where crossmodal audiovisual object stimuli have been compared to uni-modal (or intra-modal) stimuli (Beauchamp *et al.,* 2004b; Taylor *et al.*, 2006).

Beauchamp *et al.* (2004b) focused their investigation on the response properties of pSTS to uni-modal and crossmodal stimuli. Their first experiment was designed to identify regions responding to single stimuli consisting of visual pictures or environmental sounds. Regions common to both modalities were identified not only in pSTS extending into middle temporal gyrus (MTG) but also dorsolateral prefrontal cortex (DLPFC), motor cortex and ventral temporal cortex; however pSTS/MTG was the only region that responded preferentially to both modalities equivalently, relative to the baseline of scrambled visual stimuli and meaningless sounds. In their second experiment subjects carried out a semantic decision task on uni-modal auditory (environmental sound) and visual (line drawings) stimuli, and a congruency task on simultaneously presented audiovisual blocks. Three regions were identified that responded more strongly to audiovisual than the uni-modal inputs: pSTS/MTG, DLPFC and ventral temporal cortex. The experimental confound introduced by using different tasks across uni-modal and crossmodal conditions led them to carry out a further experiment, with task (semantic decision) controlled across crossmodal and uni-modal conditions, although the stimuli changed once again to moving video clips and environmental sounds. In this experiment, pSTS/MTG and ventral temporal cortex showed an enhanced response during the stimulus phase of crossmodal trials.

Throughout the set of experiments just described, pSTS/MTG responded to uni-modal auditory and visual stimuli, but when presented together the neuronal response was enhanced. This highlights two important and related questions: First, would the enhancement for crossmodal stimuli relative to the sum of two separately presented single inputs be observed if the total amount of sensory input was controlled on each trial? Only by comparing equivalent inputs, for instance by presenting uni-modal stimuli with an irrelevant control stimulus in the opposite modality, will the level of stimulus input be controlled. Second, if processing demand is equated within and between modalities (for example by engaging a task that requires matching two items for uni-modal auditory, uni-modal visual and crossmodal audiovisual trials), is an enhanced response still observed for crossmodal audiovisual trials? The data presented by Beauchamp *et al.* (2004b) showing responses in pSTS/MTG for uni-modal stimuli suggest that all trial types might show equivalent activation relative to baseline if task demands were controlled. If so, then the role of pSTS/MTG might be in the integration, or association, of two items independent of input modality. These issues are addressed in Chapter 4.

The second investigation of crossmodal conceptual processing carried out by Taylor *et al.* (2006) did control stimulus inputs and processing demand across both uni-modal and crossmodal trials. In their experiment subjects made a match/no-match decision on crossmodal trials using colour photographs presented with either environmental sounds or spoken words. Relative to uni-modal matching, where subjects either matched two halves of a photograph or two successively presented auditory sounds, they observed increased activation for crossmodal matching in bilateral medial frontal lobes, left posterior temporal cortex extending into pre-striate cortex and the left perirhinal cortex. Interestingly, activation in the medial temporal cortex interacted with congruency of matching pairs, showing a trend for increased activation for incongruent relative to congruent crossmodal pairs. This is particularly interesting in that it suggests a modulation due to semantic rather than perceptual content (i.e. two concepts are presented during incongruent trials compared with one only during congruent trials) whereas the bilateral medial frontal and left posterior temporal activations were not modulated by congruency. The role of the antero-medial temporal lobe in perceptual versus semantic processing is currently a matter of debate, as it has been found to respond to perceptual tasks with no memorial component (Bussey *et al.*, 2003; Lee *et al.*, 2005a, 2006a). This is supported by complementary data from patients with herpes simplex virus encephalitis (HSVE), who suffer predominantly atrophy to the medial temporal lobes, and can present with category-specific deficits for the category of living things (see Capitani *et al.*, 2003 for a review). A deficit for living things has been interpreted as the demands placed on perceptual differentiation of complex objects, as discussed above (Devlin *et al.*, 2002b; Moss *et al.*, 2005; Tyler *et al.*, 2004). Taken together these data highlight a need to resolve the functional role that the antero-medial temporal region plays in perceptual and/or mnemonic processing. Chapter 8 of this thesis investigates this issue in more detail.

These two studies of Beauchamp *et al.* and Taylor *et al.* raise interesting questions regarding the integration of crossmodal cues. Firstly they highlight experimental factors that require further investigation, including the modulatory effect of object congruency and the differences elicited by the variations in experimental task. Second, they implicate three regions (pSTS/MTG, ventral temporal cortex, perirhinal cortex) which appear to be differentially involved during both crossmodal and intra-modal matching. Whether these regions actively integrate crossmodal inputs is currently unclear.

I will now review data that has looked at the modulation of neural activity by object concepts from uni-modal auditory and visual inputs. This is relevant because identifying responses to multiple different input modalities within the same neural regions has been one approach for identifying candidate regions where the integration of crossmodal inputs might occur.

Functional imaging studies have consistently shown that both visual and auditory inputs access a shared semantic system (Bookheimer *et al.,* 1998; Booth *et al.,* 2002; Noppeney *et al.,* 2005, 2006; Spitsyna *et al.,* 2006; Thierry and Price, 2006), with differential activation within that system depending upon the experimental task (e.g. naming versus semantic decision) and input material (e.g. verbal versus nonverbal). Moreover, modulation of activation in response to inputs from one sensory modality has been observed in previously assumed sensory-specific processing regions for the non-preferred sensory modality (Bookheimer *et al.,* 1998; Nyberg *et al.,* 2000; von Kriegstein *et al.,* 2003; von Kriegstein and Giraud, 2006; Wheeler *et al.,* 2000). These findings make the identification of candidate sites for crossmodal integration more complex, as integration may occur at a number of levels depending upon the type of information used as stimuli or the response required for any particular experimental task. Moreover, regions shared between modalities are not necessarily regions that integrate signals from different modalities. Integration may occur at an early sensory or perceptual level prior to accessing an amodal semantic system, or it may be the result of the integration process itself within an amodal distributed system.

To summarise the literature reviewed in this section, the evidence from functional imaging studies using a single input modality shows that bilateral temporal cortices respond to visual stimulus inputs in a hierarchy shifting from posterior ventral occipito-temporal to anterior temporal lobes. Within this system, activation can be modulated depending upon the category membership of an object (e.g. natural versus manmade objects) or the experimental task (e.g. perceptual versus semantic decision), providing evidence for task- and stimulus-dependent activation within a common system. Related to this, common regions of activation have been observed for different types of material (verbal and nonverbal) and different modalities of input (auditory and visual), when task and stimulus are controlled, providing evidence for a shared system independent of modality.

The next step is to ask how crossmodal inputs are combined, and whether the integration of these crossmodal inputs occurs at different levels within this common system. That is, does semantic integration place higher demands (and hence increase neuronal activation) in the same regions identified as being engaged during semantically-mediated uni-modal tasks, or does integration occur outside of the semantic system in sensory or perceptual regions prior to accessing meaning? Moreover, does integration of perceptual cues that do not require access to meaning engage the same or different regions as those engaged during semantic tasks? Before describing the experiments in this thesis that attempt to answer these questions, the integration of tactile-visual integration will be discussed.

1.5 Integration of tactile-visual information

A search of the literature shows an interesting dichotomy between the regions purported to be involved in tactile-visual integration. Functional imaging in humans has predominantly reported on the involvement of the intra-parietal sulcus (IPS), lateral occipital lobe (LO) and insula (Amedi *et al.*, 2002; Banati *et al.*, 2000;

Grefkes *et al.*, 2002; Hadjikhani and Roland, 1998; Kawashima *et al.*, 2002; Saito *et al.*, 2003). In contrast, non-human primate studies (e.g. Cowey and Weiskrantz, 1975; Goulet and Murray, 2001; Jarvis and Ettlinger, 1977; Murray and Mishkin, 1985) and to a small, albeit inconsistent degree, neuropsychological data (Holdstock, 2005; Shaw *et al.*, 1990) have focused on the involvement of anterior medial temporal regions in successful integration of tactile and visual inputs.

One of the difficulties with a comparison across these experimental groups is that much of the functional imaging data from human subjects has focused on uni-modal tactile or visual processing; in particular the modulation of tactile inputs in visual regions such as LO (Amedi *et al.*, 2005). Only a small number of studies have directly investigated crossmodal tactile-visual integration in humans (Banati *et al.*, 2000; Hadjikhani and Roland, 1998; Saito *et al.*, 2003). Moreover, across these different human and animal studies, the experimental paradigms used engage multiple types of object integration, from perception of simple abstract shapes through to more demanding cognitive tasks that engage crossmodal recognition and mnemonic components. In the context of this thesis, investigating the integration of tactile-visual stimuli is important for 1) attempting to reconcile the contrasting findings from functional imaging studies and non-human primate data, and 2) to compare audiovisual and tactile-visual integration to investigate commonalities and/or differences between the neuronal regions involved in different types of crossmodal integration.

A number of studies have demonstrated shared object processing in the visual and tactile modalities (Amedi *et al.*, 2001; Easton *et al.*, 1997; James *et al.*, 2002; Pietrini *et al.*, 2004; Prather *et al.,* 2004). For example, Amedi *et al.* (2001) demonstrated that a lateral occipital region, which they termed LOtv, also responds when naming haptic objects in humans. Using uni-modal presentation of visual and tactile real objects and textures in a covert naming task, they found that tactile-related activation overlapped with visual activation in LOtv for objects (relative to textures). The response to both types of modality inputs for the same objects suggests that either 1) activation in LOtv reflects access to amodal or multimodal (tactile-visual) representations of objects, or 2) that visual information predominates in their naming task, and activation in response to somatosensory inputs in LOtv are the result of a reliance on visual information to perform the task, perhaps due to visual imagery. Interestingly their experimental design included an imagery control task, which did activate this occipito-temporal region, but it did so to a significantly smaller degree. In the monkey brain, the inferotemporal cortex (specifically area TE) is considered the final stage of the visual ventral pathway, and responds to moderately complex images with no variance to size or spatial differences. It has also been demonstrated by Tanaka *et al.* (1999) that neurons in this region show selective responses to grasped objects. These findings suggested that the inferior temporal cortex may be important for tactile-visual form integration. However, none of the studies explicitly measured the differences between crossmodal and intra-modal matching, and they are therefore unable to determine the level at which the two modalities are interacting, or indeed whether these signals are integrated prior to convergence in an amodal perceptual representation.

A small number of functional imaging studies in humans have been designed explicitly to investigate crossmodal relative to intra-modal tactile-visual integration.

One of the first was a PET study involving the crossmodal matching of spherical ellipsoids, carried out by Hadjikhani and Roland (1998). Specifically they were looking to identify areas responding to polysensory inputs (i.e. regions responding across tactile and visual intra-modal and tactile-visual crossmodal conditions) and crossmodal inputs (i.e. regions responding to tactile-visual matching over and above the intra-modal conditions). They did not find any areas designated as polysensory under their criteria, however the right insula/claustrum was significantly activated for crossmodal relative to intra-modal matching. Further evidence for the involvement of the insula in tactile-visual integration was provided by Banati *et al.* (2000) in their PET study, although they reported activation in the left, not right, insula. As mentioned in Section 1.3, the insula is implicated in crossmodal linguistic processing: neuroanatomical connections with auditory, visual and somatosensory systems (Pearson *et al.*, 1982) make it well-placed for the transfer of crossmodal information. In contrast, a third fMRI study by Saito *et al.* (2003) did not find activation in the insula; rather they reported only one region in the posterior IPS significant for crossmodal relative to intra-modal matching. They suggested that these differences could be accounted for by 1) task (3D shape matching in the Hadjikhani and Roland study versus 2D pattern matching in their study), 2) the experimental paradigm (Banati *et al.* did not include a tactile-tactile intra-modal condition), and 3) the imaging modality (PET versus fMRI). Clearly inconsistencies in these experimental factors make it very difficult to identify at which level in the processing of crossmodal inputs the insula or IPS are engaged, or which factors are relevant for driving activation in either of these regions.

It is perhaps surprising that ablation studies from the non-human primate literature have found that lesions to the entorhinal and perirhinal cortex impair performance on tactile-visual delayed non-matching to sample tasks (DNMS: Goulet and Murray, 2001), yet no studies have shown activation in this region when studying human subjects. In non-human primates the perirhinal cortex receives afferent projections from the adjacent uni-modal visual processing regions TE and TEO, the parahippocampal cortex, somatosensory association areas of the insula cortex, auditory association cortex in superior temporal gyrus and polymodal association cortices such as the orbitofrontal cortex and the dorsal bank of the superior temporal sulcus (Suzuki, 1996; Suzuki and Amaral, 1994a). It also has strong reciprocal connections with the hippocampus, via the entorhinal cortex, and with the amygdala (Suzuki, 1996; Suzuki and Amaral, 1994b; van Hoesen and Pandya, 1975). It is therefore well placed to combine inputs from different sensory modalities as well as to interact with other memory-related regions, integrating an object's many features into a single representation (Buckley and Gaffan, 1997; Meunier *et al.,* 1993; Murray and Bussey, 1999; Murray and Richmond, 2001). Increasing evidence from human functional imaging is supporting the involvement of the medial anterior temporal lobes, in particular perirhinal cortex, in tasks involving discrimination between complex visual featural combinations (Devlin *et al.*, 2002a; Lerner *et al.,* 2001; Moss *et al.,* 2005; Tyler *et al.,* 2004). In addition, the perirhinal cortex is activated by semantic integration of audiovisual relative to uni-modal auditory and visual signals (Taylor *et al.,* 2006). One of the questions addressed in this thesis is under what circumstances are medial anterior temporal cortices engaged by tactile-visual integration? There is an ongoing debate in the literature as to the involvement of this medial temporal region in perceptually-driven or mnemonically-driven processing

(Buckley, 2005; Buckley and Gaffan, 1997, 1998; Buckley *et al.*, 2001; Buffalo *et al.*, 1998; Bussey *et al.*, 2002, 2003; Hampton, 2005; Holdstock *et al.*, 2000; Lee *et al.*, 2005a, 2005b, 2005c, 2006a, 2006b; Levy *et al.*, 2005; Stark and Squire, 2000; Tyler *et al.*, 2004). Indeed, a relevant question is whether placing different demands on the type of processing (perceptual versus mnemonic) through the use of complementary crossmodal integration tasks (tactile-visual versus audiovisual) modulates activation. This issue is addressed in Chapters 7 and 8 of this thesis.

1.6 Overview of following chapters

Using both positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) techniques, the following set of experiments use different combinations of visual, auditory and tactile stimuli to elucidate the anatomico-functional organisation underlying object processing and representation. More specifically it asks at what anatomical and functional levels (Sensory? Perceptual? Semantic?) uni-modal and multi-modal inputs are integrated into a perceptual and/or conceptual representation. This question can be tested in a number of ways. Firstly, by manipulating (increasing) the level of perceptual information for object identification and contrasting regional activation for inputs that are within or between modality. This should identify regions where integrative mechanisms are either modality-specific or modality-independent. A second method directly contrasts the effect of crossmodal relative to intra-modal processing, in the context of the same task and stimuli. Under this scenario, activation for crossmodal processing over and above that seen for intra-modal processing would point to a crossmodal integration site, independent of a shared amodal semantic network. Third is the proposition that crossmodal integration occurs at different levels within the object processing system depending upon the cognitive operations required by the experimental task. For instance does integrating at a perceptual level versus a semantic level engage the same or a different network of regions? This can be investigated by examining the integration of crossmodal conceptual stimuli compared with meaningless abstract stimuli to determine shared and distinct regional activations.

Following an outline in Chapter 2 of the experimental methodologies used in this thesis, Chapter 3 describes a PET experiment designed to investigate the neural regions engaged in an object naming task when stimuli are manipulated by increasing perceptual cues for identification. Using both within (visual) and between (audiovisual) modality inputs of black and white photographs, colour photographs, colour patches and environmental sounds, enabled a comparison between brain regions that are engaged 1) independent of modality (visual and audiovisual), with 2) dependent on whether the additional perceptual cue is within or between modality (visual versus audiovisual). The aim was to determine whether any effect of integration occurs in brain regions associated with perceptual or semantic processing.

Chapter 4 uses fMRI to investigate the regions involved in crossmodal audiovisual integration compared with the association of the same verbal and nonverbal concepts, presented in uni-modal auditory or visual modalities. In particular, this chapter focuses on a region of interest in posterior superior temporal sulcus; an area previously associated with audiovisual integration at a perceptual, phonetic as well as semantic level. The design of the study allows the investigation of whether a task which requires the matching of audiovisual, visual or auditory stimulus pairs at a conceptual level engages different neural regions depending upon the sensory input modality.

Chapters 5 extends the findings of Chapter 4 by looking at the influence of presentation rate and congruency on audiovisual stimuli. More specifically, in Chapter 4 the use of uni-modal stimuli meant that individual pairs had to be presented sequentially. The experiment in Chapter 5 uses the same audiovisual object stimuli and matching task as used in Chapter 4, but with simultaneous presentation. The data resulting from both these experiments was combined to investigate any differences between presentation type at this conceptual level of matching.

Chapter 6 focuses on how audiovisual integration differs for verbal and nonverbal stimuli. Verbal stimuli (spoken and written words) can be matched at a phonological level whereas purely non-verbal stimuli (photographs and environmental sounds) can be matched at a semantic level. In the context of control conditions that used both verbal and nonverbal stimuli, it was hypothesised that integration of verbal versus nonverbal audiovisual pairs would differentially engage phonological and semantic processing regions respectively.

Chapter 7 moves from the integration of audiovisual conceptual stimuli to the integration of tactile-visual shape stimuli. Using the same experimental design as reported for audiovisual integration in Chapter 4, this fMRI study investigates the brain regions involved in crossmodal relative to intra-modal matching of tactile-visual abstract shapes, and whether this activation is modulated by congruency. Although there has been some consistency in the non-human primate and neuropsychological patient literature implicating the role of the anterior medial temporal lobe in tactile-visual integration, this region has not been specifically investigated in functional imaging of normal subjects.

Finally, Chapter 8 combines the crossmodal audiovisual matching data from the experiment reported in Chapter 4 and the crossmodal tactile-visual data from Chapter 7 to investigate the modulatory effect of multi-modal congruency and meaning in object processing. In particular, this comparison of perceptual versus semantic integration was ideally suited to examining the differential role that lateral versus medial anterior temporal cortex plays in object discrimination processes, based on the pattern of contrasting behavioural deficits of patients with semantic dementia (SD) and herpes simplex virus encephalitis (HSVE).

Chapter 2

2 Methods for acquisition and analysis of functional imaging data

2.1 Introduction

The functional brain imaging techniques used in this thesis include both Positron Emission Tomography (PET) and functional Magnetic Resonance Imaging (fMRI), measuring the changes in blood flow in response to increased metabolic demand as a function of neuronal activation. Imaging data obtained using these techniques are supplemented with behavioural data obtained from both within and outside the scanning environment. This chapter first gives an overview of the basic principles behind PET and fMRI measurement of the cerebral haemodynamic response, with data acquisition parameters for the experiments reported in this thesis. I will then go on to describe the data processing and statistical methods used for analysis of the obtained functional imaging data.

2.2 Positron Emission Tomography

2.21 Principles of PET

PET involves the intravenous injection of a short-lived radioactive isotope, generated in a cyclotron, used to label metabolically active molecules (Cherry and Phelps, 1995). The atoms of this isotope decay and emit a positron and a neutrino. When emitted, positrons lose energy through collision with electrons in tissue, and are very shortly annihilated. This annihilation produces gamma rays which are emitted in opposite directions, the energy of which is detected outside of the body by arrays of scintillation detectors in the PET camera. The detectors identify only coincident pairs of gamma rays on opposite sides of the head. The combinations of many lines of these responses are reconstructed by the use of a back-projection algorithm to determine the count density and position at which the collision occurred (Figure 2.1).

Figure 2.1: Signal detection in a PET camera

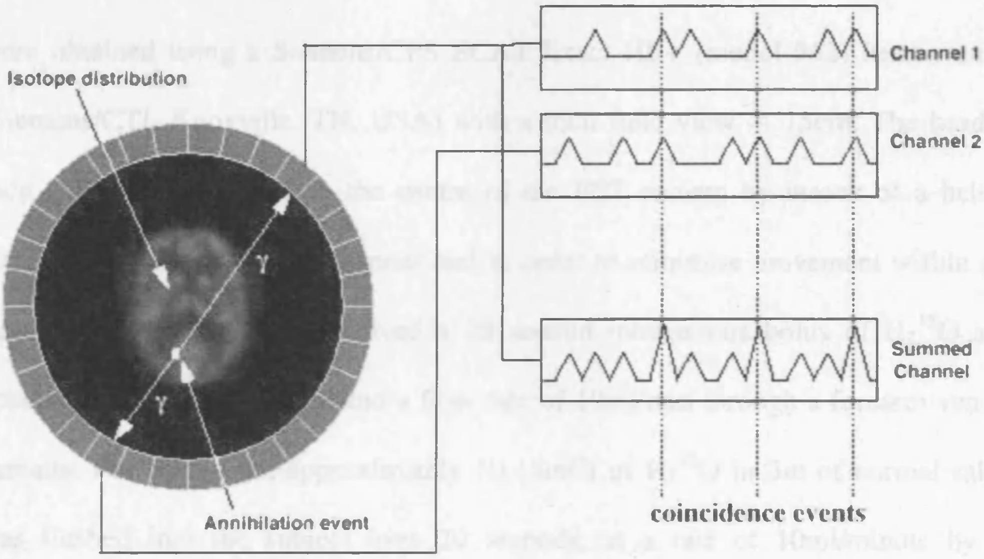


Figure illustrates gamma rays detected on opposite sides of the head by the scintillation detectors. Each detector generates a timed pulse on registering an incident photon. If pulses fall within a short time window when combined they are deemed to be coincident.

The advantage of this technique for functional brain imaging is that the particular tracer used diffuses freely across the blood-brain barrier and is taken up in cerebral tissue. Tracer uptake is therefore highly correlated with cerebral blood flow, providing a direct measurement of the cerebral haemodynamic response, which is in turn used as an indirect reference for neural activity (Petersen *et al.,* 1988; Raichle *et* *al.*, 1983). However the temporal resolution is relatively slow in order to optimise sampling of the radiotracer decay (approx 60 seconds for $H_2{}^{15}O$), and therefore requires each experimental condition to be blocked over this time period.

2.22 Data Acquisition

In the PET experiment reported here, all subjects reported underwent 12 PET relative perfusion scans at the Wellcome Trust Centre for Neuroimaging, London, UK. Scans were obtained using a Siemens/CPS ECAT Exact HR+ (model 962) head scanner (Siemens/CTI, Knoxville, TN, USA) with a total field view of 15cm. The head of each subject was located in the centre of the PET camera by means of a helmet attached with Velcro to the scanner bed in order to minimise movement within and between each scan. They received a 20 second intravenous bolus of $H_2{}^{15}O$ at a concentration of 55 MBq/ml and a flow rate of 10ml/min through a forearm venous cannula. For each scan, approximately 10-15mCi of $H_2{}^{15}O$ in 3m of normal saline was flushed into the subject over 20 seconds, at a rate of 10ml/minute by an automatic pump. After a 30 second background scan, head counts peaked 30-40 seconds later (depending on the individuals' circulation time). Data acquisition time lasted 90 seconds, with an interval of 9 minutes between successive $H_2{}^{15}O$ administrations. The assimilated radioactivity counts accumulated over the 90 second acquisition period were corrected for background noise and were used as an index of regional cerebral blood flow (rCBF).

Attenuation was corrected for by performing a transmission scan at the beginning of each study with an exposed ^{68}Ge/ ^{68}Ga external source. Images were reconstructed in 3D filtered back projection (Hanning filter, cut off frequency 0.5Hz), giving a transaxial resolution of 8.5mm full width at half maximum. The reconstructed images contained 128 x 128 pixels, each 2.05 x 2.05 x 2.00 mm in size. To ensure normal neurological status a T1-weighted structural MRI was also obtained for each participant with a Siemens Magnetom Vision 1.5T scanner (Siemens, Erlangen, Germany).

2.3 Functional Magnetic Resonance Imaging

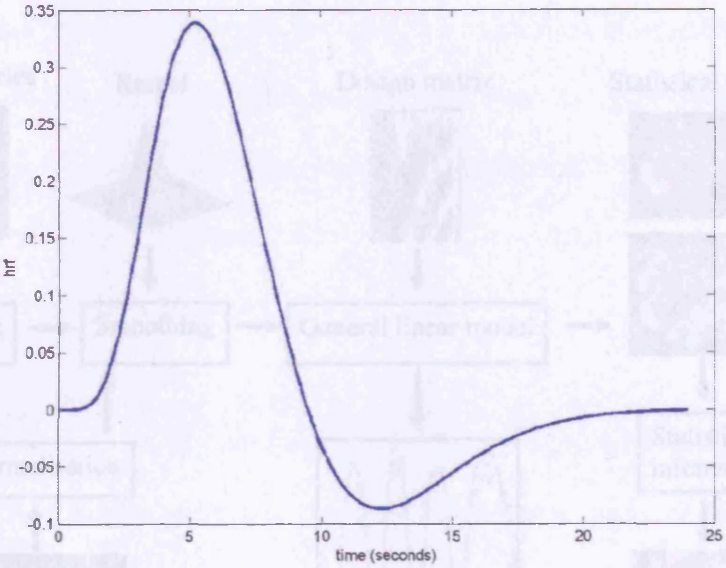
2.31 Principles of fMRI

The principles of MRI rely on the individual spinning motion of protons and neutrons within a nucleus. A hydrogen nucleus contains one positively charged proton which spins, thereby inducing a magnetic charge around it. In the absence of a magnetic field, the hydrogen spins are randomly oriented. However, when placed in a static magnetic field such as an MRI scanner, the atoms align themselves either parallel or anti-parallel to the external magnetic field ($B_{0.}$ or the longitudinal plane). A smaller number align themselves anti-parallel than parallel to the magnetic field, creating a net magnetisation parallel to the main magnetic field of the scanner (net magnetisation vector, or NMV). In addition, a secondary spin, or precession, occurs around the main axis of the magnetic field at a frequency determined by the field strength. The strength of the magnetic field used in this group of experiments is 1.5 Tesla.

When a nucleus is exposed to an external oscillating perturbation which matches its own oscillating frequency, this causes the nucleus to resonate. In MRI, a radio frequency (RF) pulse is applied that matches the frequency oscillation of hydrogen; a process known as excitation. One result of excitation is that the NMV moves out of alignment away from B_0 at an angle to the transverse plane where it generates an MR signal that can be measured using an external detector coil. When the excitation pulse ends, the net magnetisation's longitudinal component recovers at time constant T1, while the transverse component decays at time constant T2. In addition, any changes in magnetic field intensity across space will cause some spins to precess more rapidly and some more slowly, described by the time constant T2*.

Haemoglobin is diamagnetic when oxygenated but paramagnetic when deoxygenated, resulting in small differences in the MR signal of blood. Since blood oxygenation varies according to the level of neural activity, these differences can be used to detect brain activity (the blood oxygen level-dependent (BOLD) effect: Ogawa *et al.,* 1990). BOLD is thus used as an indicator of task-related neuronal activity in fMRI studies of humans (e.g., Bandettini *et al.,* 1992; Kwong *et al.,* 1992; Ogawa *et al.,* 1993). The BOLD response to an impulse stimulation (event) takes several seconds to develop and decay, with the peak haemodynamic response occurring approximately 4.7 ± 1.1 sec after a task-related change in neural activity (Aguirre *et al.,* 1998). Evoked neuronal responses are convolved with this haemodynamic response function (hrf: see Figure 2.2) to give the measured haemodynamic response (Friston *et al.,* 1994, 1995b) i.e. characterise the input-output behaviour of any voxel. This linear framework is used along with timings of experimental events and any other regressors of interest (or no interest) to model the acquired imaging data and make statistical inferences about activations in fMRI.

Figure 2.2: Function for modelling the BOLD component of the haemodynamic response



2.4 Data Analysis using Statistical Parametric Mapping

The PET and fMRI data described in this thesis were analysed using Statistical Parametric Mapping (SPM99 and SPM2, Wellcome Trust Centre for Neuroimaging, London, UK), used here to identify functionally specialised brain responses. Following data acquisition, it is necessary to carry out a number of stages of pre-processing prior to fitting the data to an appropriate experimental model. Pre-processing of the data reported in this thesis involved realignment, normalisation and smoothing. The need for implementing each of these processing steps, and an overview of the algorithms implemented in SPM2 will be described. Following that, there will be an illustration of model specification and the statistical analysis techniques used on these data. An overview of the data processing stages implemented in SPM2 is shown in Figure 2.3.

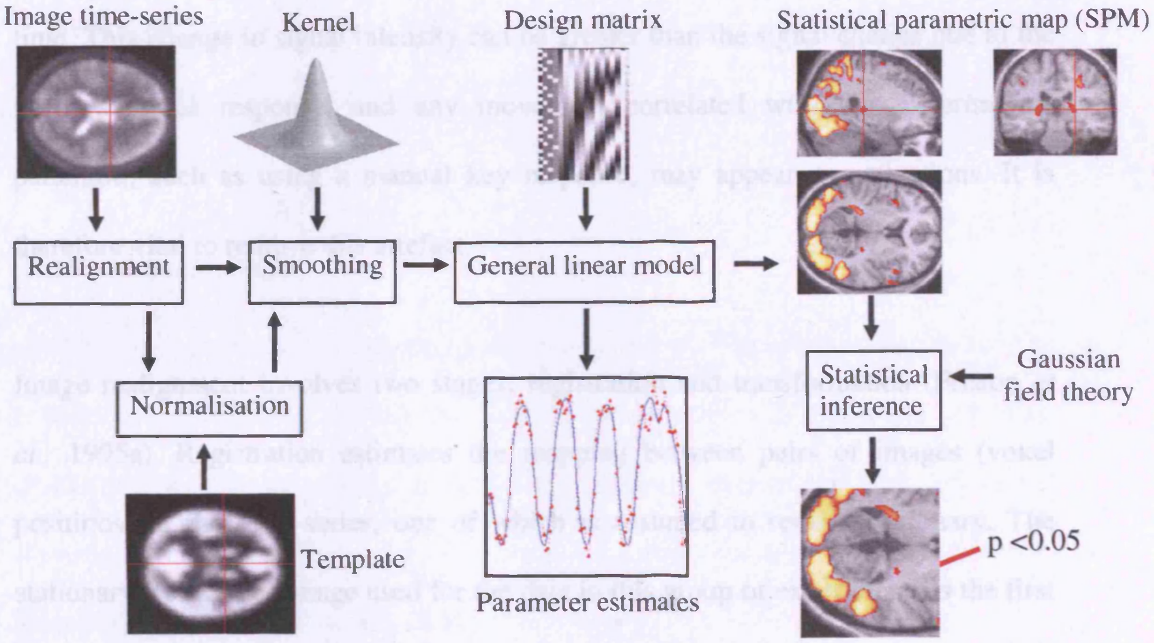


Figure 2.3: Stages of image processing and statistical analysis for PET and fMRI data, as implemented in SPM2

2.41 Pre-processing

In order to combine data from different scans both within and between subjects, images must conform to the same anatomical frame of reference. Pre-processing the data therefore involves a series of spatial transformations aimed at reducing unwanted variances induced by movement or shape differences. For the fMRI studies reported in this thesis, the first four scans of each session were removed to allow for T1 equilibriation effects prior to carrying out the pre-processing stages described here. Data were first realigned and un-warped to correct for subject movement during scanning. Second, the data were transformed using linear and nonlinear warps to normalise into a standard anatomical space. Finally, data were spatially smoothed.

2.42 Realignment

It is not possible to completely eliminate subject head movement during PET or fMRI scanning, and this movement can give rise to changes in signal intensity over time. This change in signal intensity can be greater than the signal change due to the haemodynamic response, and any movement correlated with the experimental paradigm, such as using a manual key response, may appear as activations. It is therefore vital to remove this artefact.

Image realignment involves two stages: registration and transformation (Friston *et al.,* 1995a). Registration estimates the mapping between pairs of images (voxel positions) in the time series, one of which is assumed to remain stationary. The stationary (reference) image used for the data in this group of experiments is the first image from each scanning session, after removal of dummy scans. The mapping of each voxel position to its corresponding position in the reference image is mathematically described by a set of six linear transformation parameters: three translations and three rotations about the axis. This minimises the sum of squared differences between each scan and the reference scan, thereby increasing sensitivity to true activations. These parameters are then applied to all voxels in all images, i.e. all images are transformed, or re-sampled, according to the derived parameters.

Echo-planar images contain distortions due to field inhomogeneities (i.e. susceptibility artefacts). This is problematic in the sense that subject movement interacts with this field inhomogeneity, leading to a non-rigid distortion. Because realignment uses a rigid-body model to estimate the parameters under which images are transformed, all movement may not be accounted for by the rigid-body realignment. A second process for the realignment stage ("un-warping") has been used in the pre-processing of the fMRI data presented here. This additional step uses an iterative process for estimating the movement parameters as well as the deformation fields, in order to re-estimate the movement parameters and transform all images in the time series

2.43 Normalisation

Realignment produces a mean image for each subject. This mean image is used to estimate the warping parameters that map it onto an anatomical template – an echo-planar image template which in SPM2 is the space defined by MNI (Montreal Neurological Institute, Evans *et al.,* 1992). Normalisation is carried out in a two-stage process (Andersson *et al.,* 2001).

The first stage is to estimate the transformation which minimises the (sum of squared) difference between the image and the anatomical template. The estimation determines the optimum 12-parameter affine transformation matrix to account for differences in position, orientation, and overall brain dimensions: i.e. rotation, translation, zoom and shear in each of the three dimensional planes. However, this cannot account for gross differences in head shape. Stage two uses a non-linear warp, modelled by linear combinations of discrete cosine transform basis functions. This non-linear warping is based on Bayesian inference – that is, the probability of the deformation parameters based on prior knowledge of the likely extent of deformation.

2.44 Smoothing

Following normalisation these fMRI data are then smoothed. Smoothing replaces the signal intensity of a single voxel with the weighted average of all neighbouring voxels, by applying a Gaussian kernel to each voxel (Brett *et al.,* 2004). There are four reasons for smoothing the data:

1. Smoothing the data normalises the error distribution, thereby allowing parametric testing of the effects.
2. Smoothing increases the signal to noise ratio. The optimum smoothing kernel corresponds to the size of the effect anticipated, and the spatial scale of the haemodynamic response is 2-5mm.
3. Smoothing fulfils the lattice assumption of Gaussian Random Field theory.
4. It expresses effects at a spatial scale based on homologies across different subject's functional anatomy.

The PET data reported here were smoothed at 12mm Full Width Half Maximum (FWHM). All fMRI data were smoothed at 6mm FWHM.

2.5 The General Linear Model

In order to make inferences about regional brain activity associated with experimentally induced effects, the acquired data undergoes statistical analysis, testing for effects at each cerebral voxel individually and simultaneously. This involves modelling the data into neurophysiological components of interest, confounds and error, and making inferences about the observed effects in relation to the error variance (Friston *et al.*, 1995c). The general linear model, of which t-tests, analysis of variance (ANOVA) and analysis of co-variance (ANCOVA) are variants, is an equation that expresses an observed response variable Y in terms of a linear combination of explanatory variables X plus an error term:

$$Y_j = x_{j1}\beta_1 + \ldots\ldots. + x_{jl}\beta_l + \ldots\ldots + x_{jL}\beta_L + \varepsilon_j$$

where β_l are the unknown parameters, corresponding to the L explanatory variables x_{jl}. The errors are assumed to be identically and normally distributed random variables.

The GLM can be expressed in matrix form as:

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

where Y is a column vector of observations, β the column vector of parameter estimates and ε the column vector of error terms. Matrix **X** is the design matrix which contains the explanatory variables (effects or confounds). Each column of the design matrix corresponds to an effect built into the experiment, e.g. the level of one experimental factor, or a confound. These are referred to as regressors. The relative contribution of each of these column vectors to the experimental variance (i.e. the parameter estimate β) is assessed using the maximum likelihood estimation.

2.51 Contrasts and Inference

Inferences about the contribution of the β estimates are made using their estimated variances, which allows for two types of statistical tests. The procedure results in a statistical parametric map (SPM), or 3-dimensional "image" of the statistic at each voxel. The F statistic gives an SPM{F} and tests the null hypothesis that all the estimates are zero. Alternatively, the SPM{T} tests that a particular linear combination or contrast (e.g. a subtraction of one condition from another) of the estimates is zero. The T statistic is calculated by dividing the contrast weights of parameter estimates by the standard error of that contrast. This error term is estimated using the variance of the residuals about the least square fit. For example, a contrast weight vector would be [-1 1 0 0 ....] which compares the difference in responses evoked by the first two condition-specific regressors in the design matrix.

In functional imaging analysis, the model parameters look for an effect of interest in every brain voxel. The result is a large volume of statistical values. This presents the multiple comparison problem, where the probability of reporting a false positive is increased as a function of the number of tests carried out on the voxel-wise data across the whole brain. To account for this SPM adjusts the significance (p) values in SPM based on Gaussian Random Field theory.

There are two types of inference using SPMs depending upon whether one has an *a priori* hypothesis about specific effects in a particular voxel, and this has implications for the adjustment of significance values:

1. If the hypothesis is constrained by effects relative to a particular brain region the p-value is corrected for the height or extent of that region.
2. If the hypothesis is unconstrained, the p-value must be corrected for multiple comparisons across the whole brain.

Gaussian Random Field theory takes into account the fact that neighbouring voxels are not independent by virtue of spatial smoothing in the original EPI images. It therefore controls for the expected number of false positive *regions* rather than voxels, expressing the search volume in terms of smoothness, or resolution elements (resels) of the image. Once the number of resels in an image is known, it is possible to calculate the Euler Characteristic (EC), a topological measure of the expected number of regions above any given threshold. These expected ECs can be used to give the correct threshold for the required control of false positives.

In the data presented in this thesis, unless otherwise stated, group level (random effects) activations are reported that survive a height or extent threshold of $p<0.05$, corrected for multiple comparisons across the entire brain, except in regions of interest (pSTS and anterior temporal lobe) based on *a priori* co-ordinates.

Chapter 3

3 Examining the effect of increased perceptual cues on object naming

3.1 Introduction

This chapter investigates how object naming activation is modulated by increased perceptual information. It is well documented that visual object processing proceeds in a hierarchy of stages, from early sensory input recruiting primary visual occipital cortices, through increased perceptual complexity dealt with by later stages of the visual processing stream running anteriorly along ventral aspects of the occipito-temporal lobes, and finally integration of these inputs with those previously associated with the object (e.g. motor responses and emotions) (for a review see Grill-Spector, 2003). Although a similar hierarchical model has been proposed for auditory objects (Clarke *et al.*, 2000; Maeder *et al.*, 2001; Rauschecker and Tian, 2000), it is not entirely clear how multimodal perceptual inputs are integrated into an amodal conceptual representation. Does integration occur at a semantic or a perceptual level? After a brief review of the evidence for an amodal semantic system and the effect of one type of perceptual input on the processing of another, my first experiment investigates how visual form information is integrated with colour or sound.

Data based on the observed behavioural deficits of brain damaged patients have previously led to the suggestion that semantic knowledge is stored in separable modality specific regions (Warrington and Shallice, 1984; Warrington and

McCarthy, 1994). This is based on the finding that, for example, knowledge from the auditory modality can be intact, whereas access to visual knowledge can be impaired. In contrast, distributed models of object processing have proposed an amodal semantic system, where object knowledge depends on a number of factors related to the features of the objects themselves or the processes carried out on them (Caramazza *et al.*, 1990; Lambon-Ralph *et al.*, 1999; Saffran *et al.*, 2003). The results of functional imaging studies have consistently shown that both visual and auditory inputs access a shared semantic system (Bookheimer *et al.*, 1998; Booth *et al.*, 2002; Noppeney *et al.*, 2005, 2006; Spitsyna *et al.*, 2006; Thierry and Price, 2006; Vandenberghe *et al.*, 1996). Moreover, functional imaging studies of normal subjects are now reporting effects in previously assumed sensory-specific processing regions for the non-preferred sensory modality. For example, responses in auditory cortices to visual stimuli (Nyberg *et al.*, 2000; Wheeler *et al.*, 2000), and responses in visual cortices to auditory stimuli (Bookheimer *et al.*, 1998). This suggests a sharing of information independent of modality, depending on the type of task required by the particular experimental paradigm.

Behaviourally, object naming is facilitated when perceptual cues for identification are increased. This has been found with visual stimuli, for example when objects are appropriately coloured (Humphrey *et al.*, 1994; Mapelli and Behrmann, 1997; Price and Humphreys, 1989), or crossmodal stimuli, for example when an appropriate sound is played with a visual object (Summerfield, 1992). When processing objects that share similar visual perceptual features, such as differentiating between 4-legged animals or round fruits, increased competition from structurally (or semantically) related concepts can interfere with selection of a correct representation (Humphreys *et al.,* 1995; Joseph and Gathers, 2003; McRae *et al.*, 1999). By increasing perceptual cues for identification through, for instance the provision of appropriate colour information, this competition between structurally and/or semantically related representations is decreased. Functional neuroimaging studies have associated this perceptual facilitation with differential neuronal responses. For example, Moore and Price (1999) reported a relative decrease in activation in posterior middle temporal cortices and anterior temporal cortices for natural visual objects which share perceptual features (animals and fruits) when compared with black and white line drawings of those same stimuli. In contrast to uni-modal processing, the increased stimulus input associated with crossmodal audiovisual processing *increases* activation in a network of regions, including the posterior superior temporal sulcus (pSTS; e.g. Beauchamp *et al.*, 2004b) and the antero-medial temporal lobe (e.g. Taylor *et al.*, 2006). However, see Chapters 4 and 5 for a discussion of the confounds associated with these conclusions.

The current literature therefore suggests that as the number and type of perceptual cues are increased 1) neuronal activation increases in response to the presentation of crossmodal relative to uni-modal inputs, and 2) naming activation decreases when additional perceptual cues (e.g. form and colour) are presented within a modality. These contrasting patterns of activation could either be due to task differences (passive viewing vs. naming) or fundamental differences in how multi-modal inputs are integrated within versus cross modality.

The PET experiment reported here investigates object naming from both uni-modal and crossmodal input. Specifically, within the visual modality, regions engaged when processing appropriately coloured objects (relative to their black and white counterparts and colour patches), are contrasted with activation in response to audiovisual objects (relative to their uni-modal counterparts of pictures and environmental sounds). This enables the identification of regions where object naming activation is modulated by increased perceptual cues (1) independent of whether the additional perceptual cue is within or between modality (e.g. form and colour vs. form and sound); and (2) dependent on whether the additional perceptual cue is within or between modality.

3.2 Method

3.21 Materials and Methods

Subjects

15 male subjects participated in a total of 12 x 90sec. PET scans. All were right handed native English speakers with normal or corrected to normal vision. All had normal neurological and audiological status. The study was approved by the joint ethics committee of the Institute of Neurology and University College London Hospital, London, UK.

Experimental design and stimuli

During each scan subjects were instructed to silently mouth the name of 16 objects or colour patches. The accuracy of their responses was monitored with video recording and lip reading. The stimuli were 1) photographs of objects 2) patches of colour, and

3) environmental sounds associated with the objects. Three conditions were designed to identify the effect of form (F) and colour (C), with two scans per condition. These were:

(1) Colour photos of objects with a characteristic colour (CF1)

(2) Black and white versions of condition 1 (F1)

(3) Solid colour patches (C1)

The remaining three conditions (two scans per condition) were used to investigate the integration of form (F) and sound (S):

(4) Black and white photos of objects and animals with their characteristic sounds (SF2)

(5) Black and white photographs from condition 4 without the sounds (F2)

(6) Sounds from condition 4 without the photographs (S2)

Across the Experiment, each condition was presented twice (12 scans in total). For the conditions investigating the integration of form and colour, stimuli consisted of 48 objects with a prototypical colour (24 manmade and 24 fruits and vegetables) and 12 solid colour patches. The manmade and natural objects were presented in different scans, with 12 items in the CF1 condition and the other 12 in the F1 condition. The same 12 colour patches were used in the two C1 scans. Form/sound integration conditions consisted of 36 animals and 36 manmade objects and their associated sound.

Photographs were obtained from the Hemara Photo Objects CD collection and environmental sounds were downloaded from the internet, with the majority obtained from the website www.sounddogs.com. Sounds were converted to mono and were

1500 ms. in length. The 12 colour patches were created using Corel Photo-paint v.11, and all visual stimuli were equated as far as possible for size (~8cm x 8cm). See Appendix 1 for a complete list and Figure 3.1 for examples of each trial type.

To reduce the sensory differences between uni-modal and crossmodal matching conditions, meaningful uni-modal stimuli were presented with a meaningless stimulus in the opposite modality. Thus, in the uni-modal auditory condition (S2) each environmental sound was simultaneously presented with a scrambled photograph (using the "scatter pixel" function in Corel Photo-paint v.11), and in uni-modal visual conditions (CF1, F1, C1, F2) each visual stimulus was simultaneously presented with a scrambled environmental sound. Meaningless auditory stimuli were created by converting the environmental sounds using a Fast Fourier Transform to scramble their frequency.

Stimuli were presented on a 43cm monitor suspended from a movable gantry at a distance of approximately 50cm from the subject. Sounds were presented through two speakers situated behind the subject. Stimulus presentation was controlled with COGENT software (www.vislab.ucl.ac.uk). Prior to being scanned, all subjects were familiarised with all stimuli, to ensure that they were equally familiar with the pictures and sounds. During scanning, for each condition an instruction was presented on the monitor prior to the naming task to indicate that the task was to name a particular type of object (animal, fruit, manmade object or colour). For each trial, an audiovisual stimulus was presented for 1.5s, followed by a fixation cross for 2.5s, giving an inter-stimulus interval of 4000ms and a total activation block length of 48s.

Figure 3.1 a-f: Example of a single stimulus trial for each of the six conditions

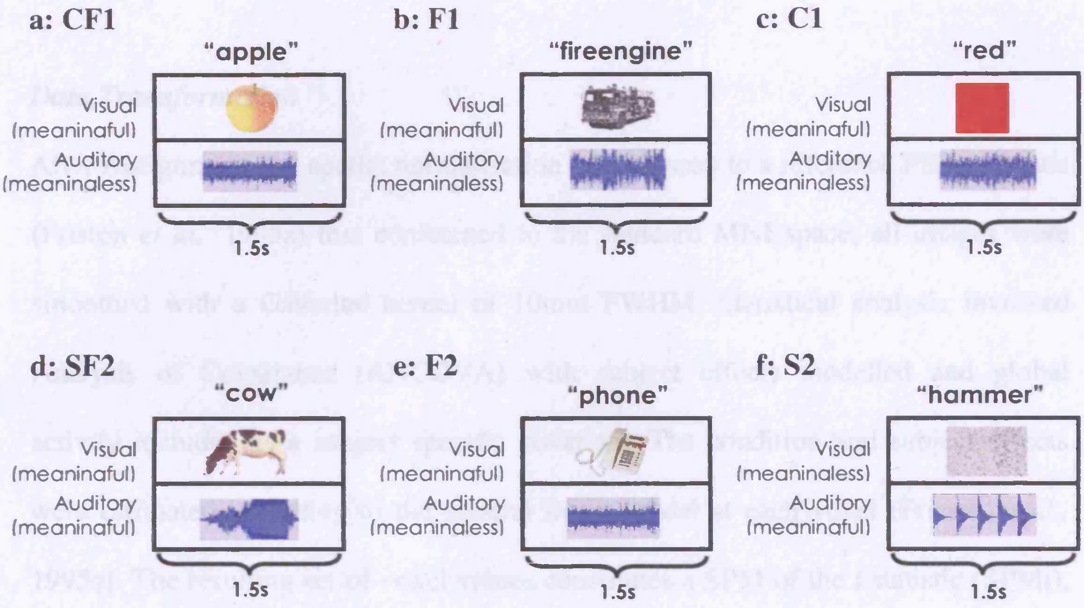


Figure shows visual stimuli consisting of pictures, colour patches or scrambled images. Auditory stimuli were environmental sounds or scrambled sounds, depicted here as the 1.5 second auditory sound envelope. Key: C = colour; F = form; S = sound; 1 = visual only; 2 = audiovisual.

3.22 Behavioural Study

As it was only possible to record accuracy but not response latencies in the scanner, an additional behavioural study was run with a separate group of 24 subjects to investigate differences in naming latencies for each condition. These subjects received the same instructions and the same 12 conditions as the previous subjects, the only difference being that they were instructed to name the stimuli out loud as quickly and as accurately as possible. Response latencies were recorded with a voice activated relay system from stimulus to response onset.

3.23 Data Acquisition

Please refer to Chapter 2.22 for PET data acquisition parameters.

Data Transformation

After realignment and spatial normalisation of each scan to a reference PET template (Friston *et al.*, 1995a) that conformed to the standard MNI space, all images were smoothed with a Gaussian kernel of 10mm FWHM. Statistical analysis involved Analysis of Covariance (ANCOVA) with subject effects modelled and global activity included as a subject specific covariate. The condition and subject effects were estimated according to the general linear model at each voxel (Friston *et al.*, 1995c). The resulting set of voxel values constitutes a SPM of the t statistic (SPMt), the values of which were transformed to the unit normal distribution (SPMZ).

In a preliminary analysis that modelled living and non-living object categories separately, the effects of additional perceptual information did not interact with category. The results reported in this chapter are therefore based on an analysis that summed over the effect of object category.

To identify regions common to naming from increased perceptual information (across sensory input type) and determine any interaction between perceptual inputs, the following contrasts were computed:

(1) Main effect of integration = [CF1> C1+F1] + [SF2> S2+F2].

(2) Interaction between form/colour integration and form/sound integration.

Greater for form/colour integration = [CF1> C1+F1] - [SF2> S2+F2].

Greater for form/sound integration = [SF2> S2+F2] - [CF1> C1+F1].

(3) Simple main effect of colour/form integration = [CF1> C1+F1]

(4) Simple main effect of sound/form integration = [SF2> S2+F2].

Unless stated otherwise, all significant effects are reported at $p<0.05$ corrected for multiple comparisons across the whole brain. There were two regions of interest (ROIs) based on previous literature. Firstly, the pSTS ROI associated with enhanced activation during audiovisual integration, which was centred on the peak co-ordinates reported in Beauchamp *et al.* (2004b) at [-50, -55, 7] for non-verbal audiovisual integration. These co-ordinates were converted from Talairach and Tournoux stereotactic space into to the nearest estimated co-ordinates in MNI space [+/-50, -56, 4] using the algorithm developed by Matthew Brett (http://www.mrc-cbu.cam.ac.uk/Imaging/Common/mnispace.shtml). The second ROI was taken from the Moore and Price (1999) study, where facilitation was observed for coloured relative to black and white natural objects. They reported bilateral effects in the antero-medial temporal region at the co-ordinates [-26, 0, -20] and [30, 8, -24]. Within these two ROIs, the search volume was a sphere of 10mm radius.

3.3 Behavioural Results

Means and standard deviations can be found in Table 3.1. For the form/colour integration, pairwise t-tests between conditions revealed that response latencies for object naming were facilitated by the presence of colour ($t=5.19$, $p<0.015$ for FC1 vs. F1) and also significantly faster when naming colour patches relative to naming both black and white pictures ($t=8.801$, $p<0.0005$) and colour pictures ($t=6.122$, $p<0.0005$). For audiovisual integration, object naming latencies were neither facilitated or inhibited when both sound and form information were simultaneously presented (*t*=0.044, p=0.965 for F2 vs. FS2) although naming objects from their sound only was significantly slower than either form only (*t*=13.062, p<0.0005) or form with sound (*t*=15.10, p<0.0005). This effect for auditory naming may be due to the presentation duration of auditory relative to visual stimuli. Auditory recognition may not occur until stimulus presentation (up to 1500ms) is complete, whereas picture presentation is complete at stimulus onset.

Table 3.1: Naming latencies (mean and standard deviation) for behavioural experiment

Naming Condition	*Mean (msec)*	*SD (msec)*
CF1	932	99
F1	966	105
C1	799	114
SF2	960	130
F2	961	162
S2	1556	273

Key: C = colour; F = form; S = sound; 1 = visual only; 2 = audiovisual

3.4 Functional Imaging Results

3.41 Whole brain analysis

Increased perceptual cues increased activation in bilateral occipital lobes and the right anterior fusiform for both the form/colour [CF1>C1+F1] and form/sound conditions [SF2>S2+F2] (see Table 3.2 and Figure 3.2). There was no significant decreased activation and no interaction between the addition of colour or sound, nor was there an interaction of these effects with category.

3.42 Regions of interest

In the ROIs, there was only one significant effect. This was in the left antero-medial temporal cortex ([-22, 8, -28; Z=3.6, p=0.023] after small volume correction centred on [-26, 0, -20]) for the interaction of form/colour and form/sound integration, see Figure 3.3. The post hoc tests show that activation was significantly less when form and colour were integrated (CF1) than for form (F1) and colour (C1) alone [-24, 8, -30; Z=3.4 for (CF1<C1+F)]. The opposite trend was observed for the form/sound condition, with a non-significant increase in activation for form and sound together relative to form and sound alone [-24, 14, -32; Z=2.3 for (SF2>S2+F2)].

Table 3.2: Regional activation for main effect of increased perceptual input

	Main effects			Z-scores for individual contrasts			
Anatomical region	**Peak cluster**	**Z-score (CF1>C1+F1) +(SF2>S2+F2)**	**Number of voxels**	**CF1>C1**	**CF1> F1**	**SF2>S2**	**SF2>F2**
R inf/mid occipital	**34, -98, -4**	6.1	740	5.4	2.9	5.3	1.7
	30, -90, -10	4.8		4.0	1.6	5.1	2.8
L inf/mid occipital	**-32, -92, -18**	5.1	581	5.5	2.7	3.9	1.8
R ant fusiform	**42, -40, -14**	4.2	239	2.8	1.8	3.8	2.3
	36 -34 -12	3.2		2.2	3.5	1.6	1.6

Table 3.2, columns 1–4, shows the anatomical regions, MNI co-ordinates and corresponding Z-scores for the main effect of increased perceptual cues across input type **(left)**. Columns 5-8 show the Z scores for the individual linear contrasts.

Figure 3.2: Main effect of increased perceptual input

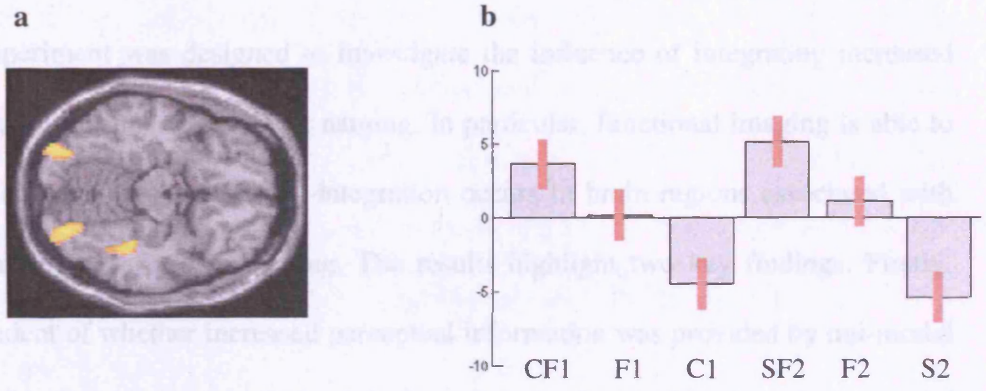


Figure 3.2a shows common activation for increased perceptual input during visual and audiovisual conditions, rendered at $p<0.001$ uncorrected on an averaged T1 section of the standardised brain at $z = -16$. 3.2b shows plot of effect size for all conditions at most significant peak in left occipital region at [34, -98, -4]

Figure 3.3: Interaction in antero-medial temporal lobe

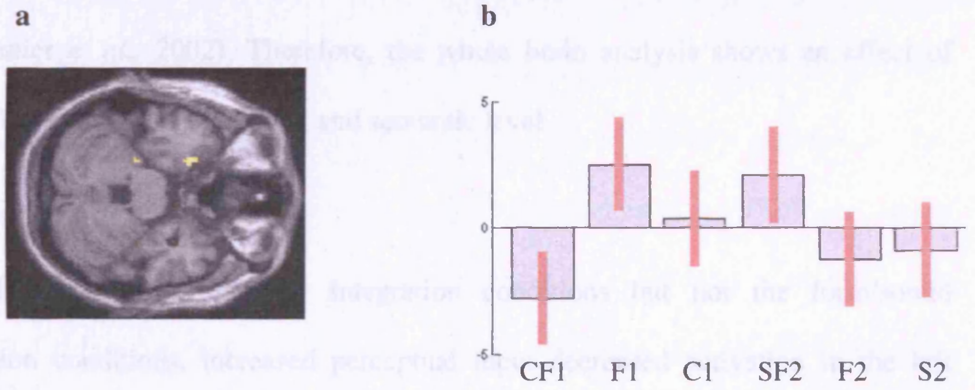


Figure 3.3a shows activation in antero-medial temporal ROI, rendered on an averaged T1 section at $z = -28$. Figure 3.3b provides corresponding plot of effect size for all conditions showing interaction at [-22, 8, -28].

3.5 Discussion

This experiment was designed to investigate the influence of integrating increased perceptual information on object naming. In particular, functional imaging is able to determine whether the effect of integration occurs in brain regions associated with perceptual or semantic processing. The results highlight two key findings. Firstly, independent of whether increased perceptual information was provided by uni-modal visual stimuli that were appropriately coloured, or by the combination of crossmodal audiovisual perceptual inputs, an enhanced response was observed in bilateral occipital cortices and the right anterior fusiform gyrus. The bilateral occipital regions are well established uni-modal visual association areas and, as such, correspond to areas involved in visual perception. In contrast, the right anterior fusiform has been associated with amodal semantic processing (Martin, 2007; Thierry and Price, 2006; Vuilleumier *et al.,* 2002). Therefore, the whole brain analysis shows an effect of integration at both the perceptual and semantic level.

Secondly, for the form/colour integration conditions but not the form/sound integration conditions, increased perceptual input decreased activation in the left antero-medial temporal lobe ROI as previously reported by Moore and Price (1999). This region has been strongly linked to semantic processing by neuropsychological studies of patients with semantic dementia and herpes simplex virus encephalitis (Barbarotto *et al.*, 1996; Brambati *et al.*, 2006; Davies *et al.,* 2004; Kapur *et al.*, 1994; Noppeney *et al.,* 2007b). The comparison of form and colour conditions to form or colour conditions (CF1 vs. F1 and C1) therefore revealed three different effects: Increased activation in areas associated with visual perception and amodal semantics; decreased activation in a different semantic region; faster response times. In contrast, the comparison of form and sound conditions to form or sound conditions (SF2 vs. F2 and S2) was only associated with increased activation in areas associated with visual perception and amodal semantics. Enhanced response for increased perceptual information will now be discussed in relation to 1) visual perceptual processing, 2) semantic processing, and 3) crossmodal audiovisual processing.

3.51 Visual perceptual processing

The posterior occipital activations associated with increased perceptual input are in well established visual processing areas (see Grill-Spector, 2003 for a review). Indeed, as shown in Table 3.2, bilateral occipital activation was most significant when object form was compared to no object form (colour only or sound only). Nevertheless, activation in these visual perceptual processing regions is also modulated by additional visual cues in the form of colour (CF1) and, more surprisingly with the addition of auditory cues (SF2). Other studies have also demonstrated modulation of visual areas by auditory inputs. For instance, Bookheimer *et al.* (1998) directly compared the semantic network involved in uni-modal auditory and uni-modal visual naming. Employing the same design as a visual object processing study (Bookheimer *et al.*, 1995), subjects were scanned during an auditory semantic task whilst blindfolded, thus ensuring no external visual stimulation. In addition to brain regions commonly associated with language processing (bilateral superior temporal, left inferior frontal, including Broca's and Wernicke's areas), activation was also observed in bilateral primary visual cortex, consistent with their previous study on visual object processing. They interpreted this as a top-down semantic to sensory effect which serves to facilitate activation of semantic representations through an automatic evocation of visual images consequent to auditory processing. Despite these previous findings, the increased occipital activation reported in this chapter for combined auditory and visual inputs remains surprising because there is no need to evoke visual imagery when the relevant visual inputs are already present. This suggests that auditory object processing automatically activates visual areas irrespective of whether visual processing is required or not. This is consistent with the observation that the addition of sound had no impact on the response times (see Table 3.1).

3.52 Semantic processing

The results show that increased perceptual input modulated activation in two different semantic regions. In the right fusiform, activation increased irrespective of whether the additional perceptual cues were colour or sound (CF1 and SF2). In the left anterior-medial temporal cortex, activation decreased when both form and colour was present but showed a trend for the opposite effect when both form and sound were present.

The fusiform has been linked to amodal semantic processing in a number of studies. For example, using a repetition priming paradigm, Vuilleumier *et al.* (2002), showed a repetition decrease in the anterior fusiform (bilaterally) for repetition of real but not nonsense pictures of objects. Martin *et al.* (1996) reported the same anterior fusiform region (also bilateral) for naming real objects relative to non-objects. Von Kriegstein *et al.* (2006) demonstrated that familiar voice and face recognition activate the right fusiform, and Thierry and Price (2006) associated right fusiform activation with amodal conceptual processing of both environmental sounds and action videos. The novel finding in the current study is that this right fusiform region shows an additive effect when both auditory and visual object inputs are presented simultaneously relative to when auditory and visual inputs are presented independently. As discussed above in relation to the corresponding effects in the bilateral occipital cortex, additional activation has no impact on object naming response times. Therefore one possibility is the right fusiform (and bilateral occipital regions) are involved in amodal processing prior to the integration of perceptual cues. Consequently, if the timing of activation depends on the input modality (e.g. faster for visual input than auditory input), then increased activation may reflect more sustained inputs when there are two sources of perceptual information (CF1 and SF2) than one (F1/F2 only or S2 only).

Although category effects have been reported in the fusiform gyri, it is worth noting that the effects observed here do not correspond to regions associated with category selective responses. For example, the right fusiform region activated here is more lateral and anterior to that reported by Chao *et al.* (1999). It is not surprising then that no interaction between category and level of perceptual information was observed.

In contrast to the effect in the right fusiform, increased perceptual information decreased activation in the left antero-medial temporal region of interest, but only for the colour/form condition not the sound form condition. This effect is consistent with that reported by Moore and Price (1999) who found increased activation for naming black and white drawings of objects relative to appropriately coloured drawings of the same objects. Because colour reduces the number of competing responses within a semantic system that shares object features, naming was facilitated, as seen in the response latencies. In contrast, there was a trend for increased activation when form and sound were presented together. If the response to perceptual information in the occipital and fusiform regions was prior to audiovisual integration, then the trend observed for increased activation in the antero-medial temporal region may be driven by semantic integration of crossmodal inputs. However, as discussed in the next section, subjects did not have to integrate both auditory and visual inputs in order to carry out the naming task, and this may have resulted in the non-significant trend observed. Further investigations of antero-medial temporal cortex are therefore required to elucidate the possible role played by this region in crossmodal processing.

3.53 Crossmodal audiovisual processing.

Given the association of pSTS with cross modal audiovisual processing (for reviews see Calvert, 2001; Calvert and Lewis, 2004; King and Calvert, 2001), it is perhaps surprising that we did not observe activation in the pSTS during audiovisual integration. In retrospect a confounding factor in the design of this study was that subjects did not have to integrate the simultaneous auditory-visual information in order to correctly respond to the task. This is reflected in the naming latencies from the behavioural experiment, where no difference was observed between the visual form only and audiovisual condition. Rather, naming was possible from the visual input alone, without recourse to the auditory signal. Attentional confounds therefore arise between crossmodal and uni-modal conditions because subjects may attend to only one modality during audiovisual conditions. The effect of attention on audiovisual processing has been demonstrated by Talsma *et al.* (2006) using ERP measurements. For example, when manipulating attention during a uni-modal auditory, uni-modal visual and bimodal audiovisual task, they found a super-additive effect for the audiovisual condition when subjects attended to both stimuli, but a sub-additive effect when subjects attended away from the objects.

3.6 Summary

To summarise, during an object naming task, combined perceptual cues in both the visual only and the audiovisual modalities increased activation in both visual perceptual (bilateral occipital) and semantic (right anterior fusiform) areas. In addition, antero-medial temporal activation was decreased for combined form/colour inputs but no areas of decreased activation were observed for the combination of visual and sound inputs. This follows the behavioural data where response times were facilitated for the combination of colour and form but not for the combination of sound and form. Increased activation in the occipital and fusiform regions for combined form and sound suggests that these effects might be arising prior to the integration of form and sound. In this case, the current experiment provides no evidence for any region involved in the integration of form and sound. One possible explanation is that the experimental design used here elicited unwanted confounds, with subjects using only the visual information for naming even within the audiovisual task. This may have been the reason why activation was not seen in pSTS – a region consistently associated with multi-sensory integration. This and other issues relating to the difficulties of design in crossmodal integration experiments are discussed in detail in the following chapters, which report a group of experiments designed to address these possible confounding factors.

Chapter 4

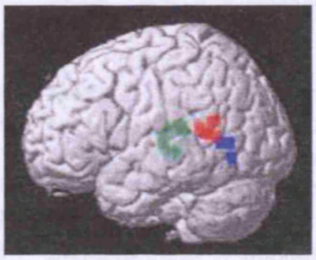
4 The role of the posterior superior temporal sulcus in auditory and visual object processing

4.1 Introduction

There has been growing interest in how the brain integrates information from different sensory modalities into a unified concept. A number of studies have associated a region of the posterior superior temporal sulcus (pSTS) with the crossmodal integration of auditory and visual information (Beauchamp *et al.*, 2004a, 2004b; Calvert *et al.*, 2000; Kreifelts *et al.*, 2007; van Atteveldt *et al.*, 2004, 2007a, 2007b; Wright *et al.*, 2003). The focus on this neural region has stemmed primarily from early neuroanatomical and electrophysiological data in non-human primates, that demonstrated the convergence of afferents from different senses within the superior temporal polysensory region – the primate homologue of human pSTS (e.g. Desimone and Ungerleider, 1986; Leinonen *et al.*, 1980; Seltzer and Pandya, 1978).

Here the evidence that a specific region in the human pSTS actively integrates auditory and visual inputs into an amodal representation is considered. According to this hypothesis human pSTS is a crossmodal "binding site" (Beauchamp *et al.*, 2004a, 2004b; Calvert *et al.*, 2000, 2001; Sekiyama *et al.*, 2003; van Atteveldt *et al.*, 2004, 2007a, 2007b). The alternative hypothesis is that pSTS activation reflects amodal processing that is independent of the sensory input modality (e.g. auditory, visual or both), for example, learnt conceptual or speech production processes that are subsequent to the processing stage where bottom up audiovisual inputs are integrated (Skipper *et al.*, 2007; van Wassenhove *et al.*, 2005). Figure 4.1 illustrates the anatomical location of the pSTS region of interest relative to more anterior STS (aSTS) sites associated with audiovisual processing and more ventral posterior middle temporal areas associated with action and tool processing. The anatomical co-ordinates of these regions are listed in Table 4.1.

Figure 4.1: Anatomical location of the region of interest



Red = areas associated with audiovisual integration in pSTS (Beauchamp *et al.*, 2004b; Calvert *et al.*, 2000; Hein *et al.*, 2007; Kreifelts *et al.*, 2007; Macaluso *et al.*, 2004; Miller and D'Esposito, 2005; Ojanen *et al.*, 2005; Olson *et al.*, 2002; Saito *et al.*, 2005; Sekiyama *et al.*, 2003; Taylor *et al.*, 2006; van Atteveldt *et al.*, 2004, 2007a, 2007b; Wright *et al.*, 2003). Green = areas associated with audiovisual integration in aSTS (Calvert *et al.*, 2000; Noppeney *et al.*, 2007a; Sestieri *et al.*, 2006; van Atteveldt *et al.*, 2004, 2007a, 2007b). Blue = areas associated with sentence, action and tool processing (Emmorey *et al.*, 2004; Kellenbach *et al.*, 2003; Lewis *et al*, 2004; Narain *et al.*, 2003; Noppeney *et al.*, 2005, 2007a; Phillips *et al.*, 2002; Vandenberghe *et al.*, 2002).

Table 4.1: Anatomical co-ordinates of temporal regions

a: Audiovisual effects in posterior STS

	Date	Visual stimuli	Auditory stimuli	Co-ordinates
Audiovisual > uni-modal				
Calvert *et al.*	2000	faces	speech streams	-49 -50 9
Wright *et al.*	2003	faces	speech (single words)	y=-40 to -55
Beauchamp *et al.*	2004b	pictures	object sounds	-50 -55 7
van Atteveldt *et al.*	2004	letters	speech sounds	-54 -48 9
van Atteveldt *et al.*	2007a	letters	speech sounds	-54 -43 13
Kreifelts *et al.*	2007	faces	speech (single words)	-54 -51 18
Saito *et al.*	2005	faces	speech sounds	*n.s.*
Taylor *et al.*	2006	pictures	sounds and words	*n.s. (-46, -76,22)*
Difficult > easy AV				
Sekiyama *et al.*	2003	faces	speech (sounds)	-56 -49 9 / -43 -55 17
Congruent > Not				
Calvert *et al.*	2000	faces	speech streams	-49 -50 9
Ojanen *et al.*	2005	faces	speech sounds	*n.s.*
van Atteveldt *et al.*	2007a	letters	speech sounds	*n.s.*
van Atteveldt *et al.*	2007b	letters	speech sounds	*n.s.*
Hein *et al.*	2007	pictures	object sounds	*n.s.*
Taylor *et al.*	2006	pictures	sounds and words	*n.s.*
Synchronous >Not				
Olson *et al.*	2002	faces	speech (words)	*n.s.*
Macaluso *et al.*	2004	faces	speech (words)	*n.s. (-64 -58 0)*
Miller and D'Esposito	2005	faces	speech sounds	*n.s.*
van Atteveldt *et al.*	2007a	letters	speech sounds	*n.s.*

n.s. = not significant

b: Audiovisual effects in anterior STS

	Date	Visual stimuli	Auditory stimuli	Co-ordinates
Audiovisual > uni-modal				
Calvert *et al.*	1999	faces	numbers (1-10)	-46 -25 13 / 57 -22 13
van Atteveldt *et al.*	2004	letters	speech sounds	-46 -19 2
Congruent > Not				
van Atteveldt *et al.*	2006	letters	speech sounds	-52 -31 15 / 60 -20 16
van Atteveldt *et al.*	2007b	letters	speech (sounds)	-47 -20 7 / -59 -33 12
Words > sounds				
Noppeney *et al.*	2007a	words/pictures	speech /sounds	-66 -27 -3
Recognition> location				
Sestieri *et al.*	2006	pictures	object sounds	-58 -18 -3

c: Semantic and sentence effects in temporal regions

	Date	Visual stimuli	Auditory stimuli	Co-ordinates
Sentences				
Narain *et al.*	2003		auditory sentences	-52 -54 14
Vandenberghe *et al.*	2002	written sentences		-52 -54 12
Tools				
Phillips *et al.*	2002	pictures and words		-58 -64 4
Noppeney *et al.*	2007a	words and pictures	auditory words	-51 -66 -6
Lewis *et al.*	2004		environmental sounds	-51 -57 3 / 49 -51 5
Actions				
Phillips *et al.*	2002	pictures and words		-58 -60 4
Kellenbach *et al.*	2003	pictures only		-48 -62 0
Emmorey *et al.*	2004	body gestures		-49 -59 0
Noppeney *et al.*	2005	written words	auditory words	-57 -63 6

The methodological principles used to identify audiovisual integration areas have been derived in the main from studies of sub-cortical structures, in particular the superior colliculus (Stein and Meredith 1993; Stein *et al.,* 1993). These principles include sensitivity to temporal and spatial correspondence, response enhancement and depression, and the rule of inverse effectiveness, where responses to crossmodal inputs are maximal when responses to individual stimuli are minimally effective. Although initial functional imaging studies of humans followed the principle that the audiovisual response should be super-additive compared to the uni-modal response (e.g. Calvert *et al.,* 2000), the relevance of this rule to fMRI data has been questioned by subsequent investigators (e.g. Beauchamp, 2005; Beauchamp *et al.*, 2004a, 2004b) who have identified audiovisual integration areas on the basis of an enhanced response to bimodal audiovisual stimuli relative to either auditory or visual stimuli alone.

In this and the following Chapter, the role of the pSTS is investigated by examining its response characteristics. If the pSTS actively binds auditory and visual inputs then I would expect the response to audiovisual inputs to be greater than the response to uni-modal inputs (Beauchamp *et al.*, 2004b, 2005; Calvert *et al.*, 2000; van Atteveldt *et al.,* 2004). In contrast, if pSTS activation reflects amodal processing that is independent of the stimulus modality, then I would expect the response to depend on the task demands but not the stimulus modality.

Early perceptual integration has been investigated with stimuli such as tones and circles that have no *a priori* relationship (e.g. Bushara *et al.,* 2001; Degerman *et al.*, 2007; Giard and Peronnet, 1999). This differs from phonetic or conceptual stimuli that involve top-down processing from prior knowledge (Skipper *et al.,* 2007; van Wassenhove *et al.,* 2005). In studies of continuous and meaningful speech (e.g. Calvert *et al.,* 2000), audiovisual information converge at both a phonetic and conceptual level. Phonetic without conceptual convergence can be studied using temporally brief speech sounds (e.g. "ta" or "ba") that are heard while viewing mouths articulating the same or different sounds (Miller and D'Esposito, 2005; Ojanen *et al.,* 2005; Saito *et al.*, 2005; Sekiyama *et al.*, 2003; Skipper *et al.*, 2007) or written letters presented with their auditory speech sounds (van Atteveldt *et al.*, 2004, 2007a, 2007b). In contrast, conceptual without phonetic convergence can be studied using pictures or objects and their associated auditory sounds (Beauchamp *et al.*, 2004b; Hein *et al.*, 2007; Sestieri *et al.*, 2006; Taylor *et al.*, 2006).

The initial claims for pSTS as an audiovisual integration site came from Calvert and colleagues (2000) who contrasted audiovisual speech to each modality in isolation (i.e. heard words or silent lip-reading). This revealed a super-additive response in left pSTS when the audiovisual input was congruent but a sub-additive response when the audiovisual input was incongruent. Subsequently, pSTS activation has been associated with bimodal audiovisual stimuli in studies using single words (Kreifelts *et al.*, 2007; Wright *et al.*, 2003), phonetic stimuli (Sekiyama *et al.*, 2003; van Atteveldt *et al.*, 2004; 2007b) and conceptual stimuli (Beauchamp *et al.*, 2004b). However, there are two points of inconsistency. The first is that other studies (e.g. Saito *et al.*, 2005; Taylor *et al.*, 2006) did not report enhanced activation in our pSTS region of interest for bimodal audiovisual relative to uni-modal stimuli. The second is that the super- and sub-additive effects in the pSTS for congruent and incongruent bimodal stimuli (Calvert *et al.*, 2000) have not been replicated, see Table 4.1a for a summary.

Below I reconsider the evidence that has been used to associate human pSTS with audiovisual integration. This evidence includes enhanced responses for 1) audiovisual relative to auditory or visual inputs alone, and 2) congruent relative to incongruent audiovisual stimuli. On the basis of this review, it is argued that the current data do not support rejecting the alternative hypothesis: i.e. that pSTS activation is not selective for audiovisual processing, but instead reflects amodal processing that is independent of stimulus modality. In this context, I will now highlight criteria that need to be controlled in fMRI studies of audiovisual integration.

4.11 Bimodal versus uni-modal stimuli

The association of pSTS with audiovisual integration has been supported by observations that the response to bimodal audiovisual stimuli is greater than the response to the uni-modal parts. The problem here is that if the bimodal response is not super-additive, then there is no clear way to distinguish the integration process from subsequent (downstream) amodal processing. This is because both explanations predict higher activation when there are two stimuli (e.g. bimodal) than when there is only one stimulus (i.e. uni-modal). This was demonstrated in several early functional imaging experiments that observed linear and nonlinear activation increases in perceptual and semantic processing regions when the stimulus input rate increased (Fox, 1989; Price *et al.,* 1992, 1996; Binder *et al.,* 1994). The interpretation of enhanced or super-additive activation for bimodal relative to uni-modal stimuli therefore depends on whether the total stimulus input has been controlled on a trial by trial basis, for example by comparing bimodal audiovisual stimuli to trials that present two uni-modal visual stimuli or two uni-modal auditory stimuli.

In the review of the literature on audiovisual phonetic and conceptual processing (see Table 4.1), the only studies that compared bimodal audiovisual stimuli to dual presentation of two auditory or visual stimuli were the only studies that did not observe enhanced activation in our pSTS region of interest (Saito *et al.,* 2005; Taylor *et al.,* 2006). Instead, Saito *et al.* (2005) discuss the importance of parietal regions and Taylor *et al.* (2006) discuss the importance of the perirhinal cortex and an occipito-temporal area that lies 2.5cm posterior to the pSTS region of interest (see Table 4.1a). Put another way, none of the experiments reporting pSTS activation for audiovisual stimuli controlled for the total stimulus input per trial. Instead, they compared bimodal stimuli with trials that only present one single uni-modal stimulus.

4.12 Effect of task

It has already been shown in the experiment described in Chapter 3 that the naming task used did not lead to activation in the pSTS, and this might have been because subjects did not use both auditory and visual information to respond. The comparison of a bimodal stimulus to a single uni-modal stimulus makes it difficult to find experimental tasks appropriate to both types of presentation. Indeed, as reviewed above, most previous experiments comparing bimodal to uni-modal processing have not controlled for the total stimulus input per trial. In addition to stimulus confounds, some of the studies reporting pSTS activation for audiovisual stimuli were also confounded by task differences. For example, Beauchamp *et al.* (2004b; Experiment 2) compared a same/different matching task on audiovisual stimuli with a semantic decision task on the uni-modal stimuli (e.g. four or two legs? with true/false response). Increased activation in left pSTS during audiovisual matching may therefore have been driven by task as well as stimulus rate differences. Likewise, when no explicit task is required (Calvert *et al.*, 2000, 2001; Wright *et al.*, 2003; van Atteveldt *et al.*, 2004; 2007a) or during the one-back task (Beauchamp *et al.*, 2004b), attentional confounds can arise because subjects may focus on one modality only during bimodal trials. To date, the only study that matched total stimulus input and task when comparing bimodal to uni-modal conceptual matching (Taylor *et al.*, 2006) did not observe activation in the pSTS region reported in other fMRI studies of audiovisual integration.

4.13 Congruent versus incongruent matching

The effect of congruency on audiovisual processing in the pSTS was highlighted by Calvert *et al.* (2000) who observed a super-additive effect for congruent audiovisual stimuli (AV>A+V) and a sub-additive effect of incongruent audiovisual stimuli (AV<A+V). An effect of audiovisual congruency has also been reported by van Atteveldt *et al.* (2007a, 2007b) but their effects were located in aSTS, not pSTS (see Table 4.1b). In fact, I was unable to find any studies that have replicated the Calvert *et al.* (2000) study showing enhanced pSTS activation for congruent relative to incongruent bimodal stimuli.

There are three possible reasons to explain why the effect of congruency in pSTS as reported by Calvert *et al.* (2000) has not been replicated. The first is that the stimuli used by Calvert *et al.* (2000) were speech streams (stories) whereas the other studies comparing congruent and incongruent stimuli used temporally brief stimuli (e.g. single speech sounds, words or environmental sounds) that did not offer a sufficiently long time frame to enhance audiovisual integration (see Calvert and Lewis, 2004). The problem with this explanation is it that it does not explain why pSTS has been associated with audiovisual integration in studies using temporally brief sounds and words (Beauchamp *et al.,* 2004b; Kreifelts *et al.,* 2007; van Atteveldt *et al.,* 2004, 2007a; Wright *et al.,* 2003).

The second (but not mutually exclusive) explanation is that the congruency effects reported in Calvert *et al.* (2000) arose from stimulus interference effects. For example, when audiovisual speech streams are congruent, speech comprehension benefits from both auditory and visual processing. In contrast, when audiovisual inputs are incongruent, speech comprehension is impaired because the visual information conflicts with the auditory information (see Calvert and Lewis, 2004). Reduced pSTS activation for incongruent speech streams may therefore reflect reduced comprehension. Although this explanation is consistent with studies showing enhanced pSTS activation for written and auditory sentence comprehension (Narain *et al.,* 2003; Vandenberghe *et al.,* 2002, see Table 4.1c), it does not explain why pSTS activation is activated by non-semantic speech stimuli that have no meaning or syntax (Sekiyama *et al.,* 2003).

The third explanation for inconsistent congruency effects arises from previous observations that attention to one modality only during bimodal presentation elicits sub-additive effects (Talsma *et al.,* 2007; Talsma and Woldorff, 2005). It is therefore possible that, to minimise interference during incongruent audiovisual speech streams, subjects may automatically or attentionally reduce visual processing (Deneve and Pouget, 2004; Ernst and Bulthoff, 2004) particularly in the study by Calvert *et al.* (2000) where congruent and incongruent conditions were presented in separate experiments with no instructions to attend to the visual stimuli. This would explain the absence of congruency effects in studies that presented brief stimuli or forced subjects to attend to the visual input during incongruent audiovisual conditions.

This Study

To summarise, several studies have highlighted pSTS as an area that is important for the integration of audiovisual conceptual and phonetic inputs even when temporally brief phonetic or conceptual stimuli are used. However, none of these studies equated the number of stimuli per trial or observed an effect of temporal synchrony on the audiovisual stimuli. Moreover, only one study observed super- and sub-additive effects of audiovisual congruency and this was during a passive task that was susceptible to audiovisual interference effects in the incongruent condition. The role of this pSTS region therefore requires further investigation.

The present study investigates the role of the pSTS during audiovisual conceptual processing, while keeping the task and number of stimuli constant. To address the task, stimulus and attentional confounds outlined above, task and stimulus input remained constant across conditions by manipulating the sensory modality of object pairs. These pairs comprised either audiovisual stimuli (one picture and one spoken name, or one environmental sound and one written name) or intra-modal stimuli in either the visual or auditory modalities (one picture and one written name vs. one environmental sound and one spoken name). By presenting stimulus pairs in each condition (intra-modal and crossmodal), the overall stimulus input was held constant. Moreover, the use of a within-pair match/no-match decision ensured that subjects had to attend to both stimuli in order to carry out the experimental task. An additional experiment was also carried out, in which subjects were presented with pairs of meaningless audiovisual stimuli. The data derived from this experiment was used in order to exclude activation related to meaningless sensorimotor processing.

4.2 Method

4.21 Materials and Method

Subjects

There were 18 subjects in Experiment 1a (11 women, 7 men, mean age 26), and 8 subjects in Experiment 1b (4 women, 4 men, mean age 30). All were right handed native English speakers with normal or corrected to normal vision. All had normal neurological and audiological status. The study was approved by the joint ethics committee of the Institute of Neurology and University College London Hospital, London, UK.

Experiment 1a: Bimodal relative to uni-modal object matching

In this experiment, subjects were instructed to indicate, via a left or right hand key pad response, whether two successively presented stimuli referred to the same object or not. There were four types of object stimuli: pictures of objects, their written names, their auditory names and their associated environmental sounds (i.e. two visual and two auditory). To avoid differences in divided attention within and across modality, presentation of stimuli within a pair was sequential with no inter-stimulus interval (i.e. the onset of the second stimulus corresponded to the offset of the first stimulus). In total there were 8 different stimulus conditions that presented the same stimuli in different combinations:

Each crossmodal trial had either:

(1) One photograph followed by one spoken name,

(2) One spoken name followed by one photograph,

(3) One written name followed by one environmental sound,

(4) One environmental sound followed by one written name.

Each intra-modal trial had either:

(5) One photograph followed by one written name,

(6) One written name followed by one photograph,

(7) One spoken object name followed by one environmental sound,

(8) One environmental sound followed by spoken object name.

The contrast of interest was the difference between activation for the two crossmodal conditions and the two intra-modal conditions. This entailed a manipulation of the stimulus pairings while the stimulus subcomponents were held constant. The order of the subcomponents (e.g. photograph followed by written name or written name followed by photograph) resulted in two versions of each condition type. In addition, the subcomponents were either congruent (both referred to the same object, requiring a Yes-they match response) or incongruent (referring to two different objects and requiring a No-they do not match response). This resulted in a total of 16 different trial types. The effect of congruency was randomised but the 8 types of stimulus combination were blocked with 3 congruent and 3 incongruent trials per block. Trial duration was 3.24s (1s for each stimulus followed by 1.24s fixation to allow for the response). This resulted in a total block time of 19.44s. Blocks were followed by a period of fixation which alternated between 2.7s and 13.5s. Over the experiment, each subject was presented with 9 blocks of each stimulus combination with a total of 72 blocks split into 4 different scanning sessions. Within each session, there were 6 blocks of crossmodal audiovisual matching, 6 blocks of intra-modal visual matching and 6 blocks of intra-modal auditory matching. Over sessions, the order of conditions was counterbalanced within and between subjects. Stimuli within a block were always from the same object category and the three object categories were fully counterbalanced across conditions, sessions and subjects. However, as predicted on the basis of Figure 4.1, there were no differential effects of category in the pSTS region of interest, therefore in the analyses described below, I sum over the effect of category.

Stimuli

The four types of object stimuli (colour photographs, written objects names, spoken object names and environmental sounds) all referred to the same set of 108 animals and manmade items. See Appendix 2 for a complete list. Photographs were obtained from the Hemara Photo Objects CD collection, environmental sounds were downloaded from the internet, with the majority obtained from the website www.sounddogs.com, spoken words were recorded by a female English speaker in a sound proof room and written words were presented in arial font. Note that the physical characteristics of the stimuli were controlled in crossmodal and uni-modal conditions because the individual stimuli were held constant while the pairings of these stimuli were varied. Visual pictorial stimuli were presented using a rear projector viewed via a mirror mounted on the head coil and equated as far as possible for size (~10cm x 10cm). Written words subtended a viewing angle of 1.0 - 4.0° (width) x 0.72° (height). All sounds were presented in mono via MRI-compatible electrostatic headphones and normalised using a low-pass 4th order Butterworth filter at 5000 Hz. All stimuli were 1000ms in duration, except for spoken words which had a range of 650ms – 1000ms.

The use of uni-modal and sequential stimuli is subject to gross differences in visual and auditory attention across conditions. In an attempt to reduce these differences, each object stimulus (i.e. a photograph, word or sound referring to an object) was presented with a meaningless stimulus in the opposite modality. Thus, each environmental sound was presented with a scrambled photograph (using the "scatter pixel" function in Corel Photo-paint v.11) that removed all recognisable structure. Each spoken word was presented with a row of XXXs (matched to the number of letters in the corresponding written word). Each photograph was simultaneously presented with a scrambled environmental sound and each written name was simultaneously presented with a scrambled spoken word. The scrambled auditory stimuli were created by converting the environmental sounds and spoken words using a Fast Fourier Transform to scramble their frequency. This resulted in meaningless auditory stimuli which sounded like white noise with no phonetic content. See Figure 4.2 for schematic of each experimental trial type.

Experiment 1b: Baseline task for sequential matching

To reduce the impact of non-conceptual sensorimotor processing further, a further baseline condition was created. This involved only the meaningless audiovisual stimuli from Experiment 1a, therefore it was not possible to use the object matching task. Instead, subjects were instructed to make an alternating key-press response (right-left) at the end of the second stimulus. Data for this subsidiary experiment was conducted on a different day from Experiment 1a.

The pairings of the four different meaningless stimulus types resulted in 4 different conditions:

(1) Scrambled photograph followed by scrambled spoken word,

(2) Scrambled spoken word followed by scrambled photograph,

(3) Row of XXXs followed by a scrambled environmental sound,

(4) Scrambled environmental sound followed by row of XXXs.

The timing of stimulus presentation, and fixation was identical to Experiment 1a (see Figure 4.2). With half the number of conditions (4 in Experiment 1b and 8 in Experiment 1a), it was only necessary for each subject to participate in two scanning sessions (as opposed to four in Experiment 1a). The number of trials per condition was therefore held constant across Experiments.

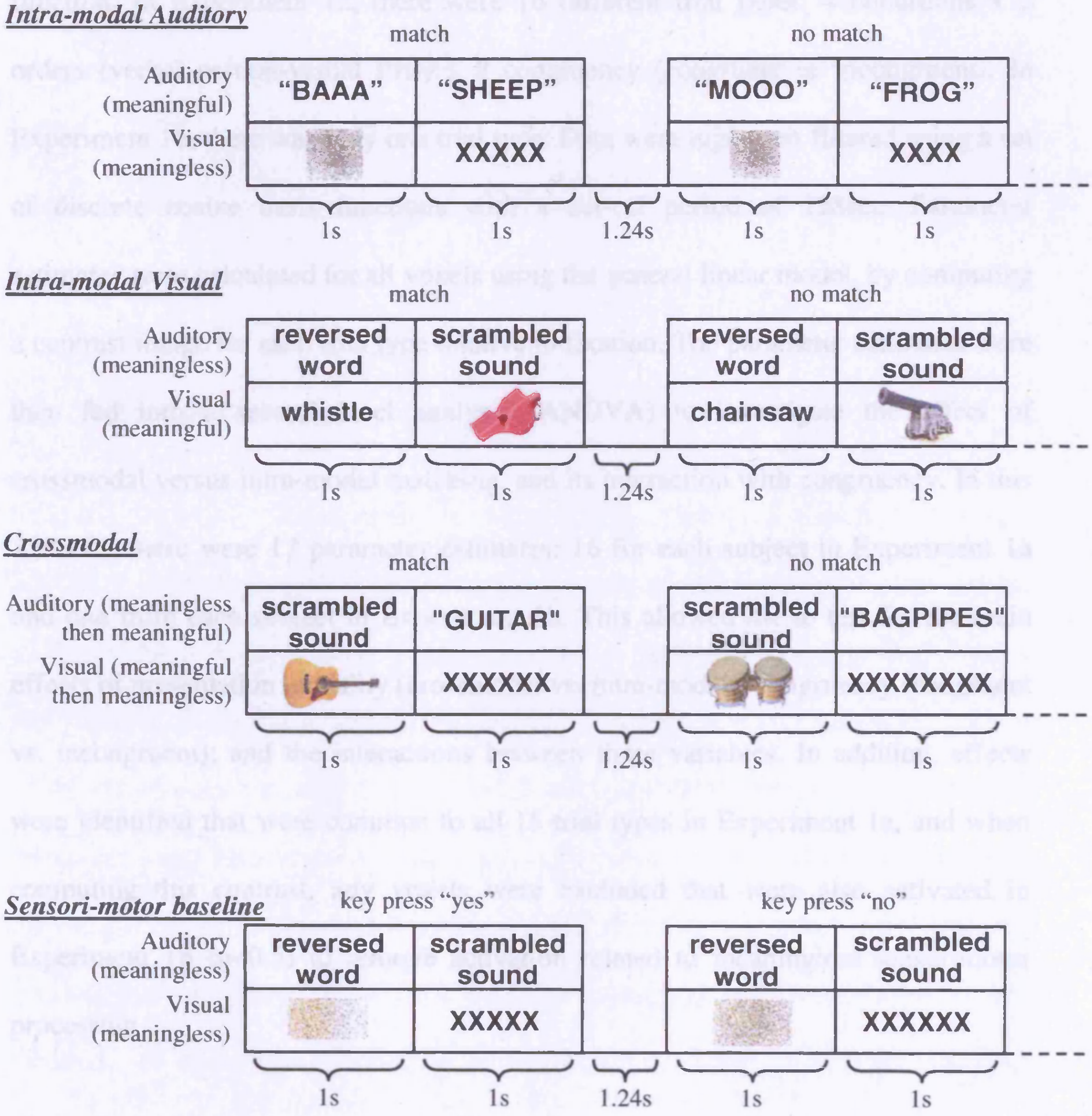
4.22 Data Acquisition

Data were acquired on a Siemens 1.5 Tesla scanner (Siemens, Erlangen, Germany). Functional images used a T2*-weighted echo-planar (EPI) sequence for BOLD contrast with 3 x 3 mm in plane resolution, 2mm slice thickness and a 1mm slice interval. 30 slices were collected, resulting in an effective repetition time (TR) of 2.7 sec/volume. After the functional sessions, a T1-weighted anatomical volume image was acquired from all subjects to ensure normal neurological status.

Data Analysis

Functional data were analysed with statistical parametric mapping (SPM2, Wellcome Trust Centre for Neuroimaging, London, UK) implemented in Matlab 7.1 (Mathworks, Sherborne, MA, USA). Pre-processing included realignment and unwarping using the first volume as the reference scan (after excluding the first 4 dummy scans to allow for T1 equilibration effects) spatial normalisation to a standard MNI template (Friston *et al.,* 1995a) and spatial smoothing using a 6mm full width half maximum isotropic Gaussian kernel. One subject was removed from the analysis due to excess head movement.

Figure 4.2: Stimulus trials for Experiments 1a and 1b



In Experiments 1a and 1b each trial (6 per block) consisted of two simultaneously presented audio and visual stimuli, followed immediately by two further audio and visual stimuli with a fixation cross presented between trials. In Experiment 1a, a key press response was made as soon as subjects could determine whether the stimuli referred to the same concept or not. In Experiment 1b subjects alternated a left and right key press response.

First level statistical analyses (single subject and fixed effects) modelled each trial type independently by convolving the onset times with the haemodynamic response function. In Experiment 1a, there were 16 different trial types: 4 conditions x 2 orders (verbal or non-verbal first) x 2 congruency (congruent or incongruent). In Experiment 1b, there was only one trial type. Data were high-pass filtered using a set of discrete cosine basis functions with a cut-off period of 128sec. Parameter estimates were calculated for all voxels using the general linear model, by computing a contrast image for each trial type relative to fixation. The parameter estimates were then fed into a second level analysis (ANOVA) to investigate the effect of crossmodal versus intra-modal matching, and its interaction with congruency. In this ANOVA there were 17 parameter estimates: 16 for each subject in Experiment 1a and one from each subject in Experiment 1b. This allowed me to test for the main effects of presentation modality (crossmodal vs. intra-modal); congruency (congruent vs. incongruent); and the interactions between these variables. In addition, effects were identified that were common to all 16 trial types in Experiment 1a, and when computing this contrast, any voxels were excluded that were also activated in Experiment 1b ($p<0.5$) to remove activation related to meaningless sensorimotor processing.

Statistical Threshold

The t-images for each contrast at the second level were subsequently transformed into the statistical parametric maps of the Z statistic. Unless stated otherwise, all significant effects are reported at $p<0.05$ corrected for multiple comparisons either across the whole brain or in the pSTS ROI which was centred on the peak co-ordinates reported in Calvert *et al.* (2000: -49, -50, 9) and Beauchamp *et al.* (2004b:

-50, -55, 7); for verbal and nonverbal audiovisual integration respectively. These co-ordinates were then converted from Talairach and Tournoux stereotactic space into to the nearest estimated co-ordinates in MNI space using the algorithm developed by Matthew Brett (http://www.mrc-cbu.cam.ac.uk/Imaging/Common/mnispace.shtml). Within these two ROIs (transformed to [+/-50, -52, 8] and [+/-50, -56, 4]), a sphere with a 6mm radius was searched for the nearest peaks in this data set.

4.3 Behavioural Results

Reaction times for Experiment 1a were analysed using a repeated measures ANOVA, modelling presentation modality (crossmodal audiovisual, intra-modal auditory and intra-modal visual) and congruency (congruent, incongruent). Means and standard deviation are shown in Table 4.2. A 3x2 ANOVA identified a main effect for sensory modality ($F[2,18] = 22.968$, $p<0.0005$), but no main effect for congruency and no interaction. Pairwise comparisons across modality revealed that response latencies increased from visual-visual (VV) to audiovisual (AV) to auditory-auditory (AA) trials. This response pattern is not consistent with that observed in early sensory integration experiments, where a bimodal stimulus facilitates a task response relative to a uni-modal stimulus. However, Beauchamp *et al.* (2004b, Experiment 3) also report increases in response time from V to AV to A conditions. The most likely explanation is in terms of the difference in duration of auditory relative to visual stimuli. Visual matching is fastest because subjects can make their decision at the onset of the stimulus. Auditory matching is slowest because stimulus recognition may not occur until stimulus presentation is complete (up to 1000ms). During AV matching, reaction times are slower than visual but faster than auditory because half the second stimuli are auditory and the other half are visual.

Table 4.2: Behavioural data

Matching Condition	*Mean (ms)*	*SD*
AV con	1001	214
AV inc	1029	211
VV con	927	246
VV inc	927	250
AA con	1114	166
AA inc	1122	168

Mean and standard deviation for reaction times in response to audiovisual, visual and auditory matching tasks in Experiments 1a. Data are for 10 subjects (due to technical difficulties with recording from the keypad). Key: con = congruent trials, inc = incongruent trials, AV = audiovisual matching, VV = visual matching, AA = auditory matching.

4.4 Functional imaging analysis

4.41 Intra-modal matching

As expected, matching visual pairs activated bilateral visual cortices relative to the AV and AA conditions, and auditory pairs activated auditory cortices relative to AV and VV (see Figure 4.3).

Figure 4.3: Main effect of intra-modal matching

Rendered on the SPM averaged MNI surface template, with VV > AV +AA in red and AA > AV + VV in green, at a threshold of p<0.001 uncorrected.

4.42 Crossmodal versus Intra-modal matching

There was no significant difference between the crossmodal conditions (AV) and the mean response of the intra-modal (VV and AA) conditions (AV > mean[VV+AA]) for either congruent, incongruent or the sum of both congruent and incongruent. Even when the threshold was reduced to $p<0.05$ uncorrected in the left and right pSTS ROIs, no voxels were identified with increased activation for AV > mean[VV+AA]. This is because activation in the pSTS ROIs was part of a widely distributed system that was activated for intra-modal as well as crossmodal matching. Figure 4.4 shows the activation pattern for each condition in Experiment 1a relative to fixation before (a) and after (b) the sensorimotor areas activated in Experiment 1b have been removed. The remarkable consistency in the location of the activation peaks for crossmodal and intra-modal matching relative to fixation is demonstrated in Appendix 3. The peak co-ordinates in the pSTS ROIs, after sensorimotor activation had been removed, were identified at [-50, -50, 10/54, -54, 8] for AV and VV and [-50, -50, 10/54, -54, 6] for AA. These effects are within 2mm of the centre of the regions of interest [+/-50, -52, 8] and [+/-50, -56, 4]) based on previous studies of audiovisual integration (see Methods). The Z scores associated with the pSTS effects were also highly significant and greater in the left than right hemisphere (Z=Inf/3.5 for AV; 5.8/3.2 for VV; and Inf/3.8 for AA).

4.43 Effect of Congruency

Congruent relative to incongruent activation was observed in bilateral caudate/thalami [6, 14, 2, Z=4.98; 16, 4, 10; Z=4.64; -10, -2, 10, z=4.2]. This was the result of decreased activation for all incongruent trials relative to fixation. There was no significant effect of incongruent relative to congruent during any condition, and no effect of congruency or incongruency in the pSTS ROI ($p<0.05$ uncorrected).

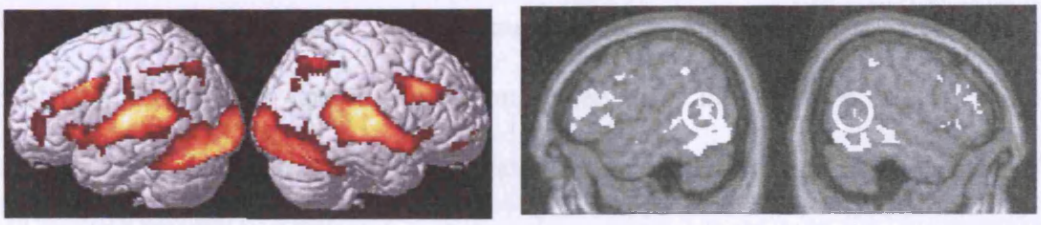
4.44 Summary

Experiment 1a identified a bilateral network of brain regions involved in the association of two meaningful stimuli, and determined that no region within this network was activated by crossmodal more than intra-modal matching, and no effect of congruency. This common network included the pSTS region previously implicated in multi-sensory integration. Even when activation related to sensori-motor processing was removed, activation in the left pSTS remained across all conditions.

Figure 4.4: A common system for matching crossmodal and intra-modal stimuli

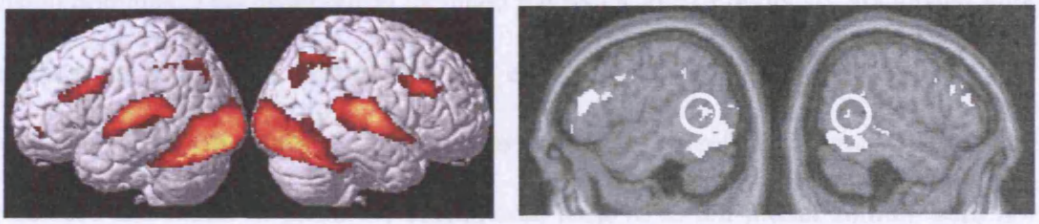
Audiovisual matching

a: minus fixation **b. minus sensorimotor baseline**



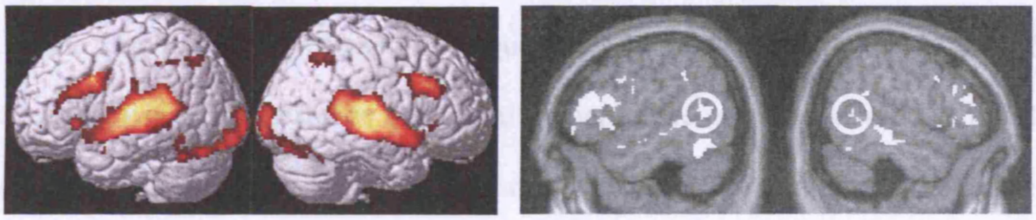
Visual-visual matching

a: minus fixation **b. minus sensorimotor baseline**



Audio-audio matching

a: minus fixation **b. minus sensorimotor baseline**



Increased activation for sequential matching relative to fixation before (a) and after (b) removing sensorimotor activation in Experiment 1b. The statistical threshold was $p<0.05$ corrected for multiple comparisons when the baseline was fixation (a) and $p<0.01$ uncorrected on saggital sections (x=-50 on left, x = 54 on right) when the baseline removed sensorimotor processing (b). A white circle highlights pSTS, activation. Across all three conditions, the peak co-ordinate in left pSTS was centred at [-50, -50, 10]. In the right, the peak co-ordinate was centred for AV and VV at [54, -54, 8] and for AA at [54, -54, 6].

4.5 Discussion

This study highlights three findings that have implications for our understanding of the conceptual association of object stimuli in the auditory and visual modalities. First, the Experiment demonstrated that, when task and stimulus factors were controlled, the conceptual network activated by sequential audiovisual matching of object stimuli was equally activated by intra-modal matching in the auditory or visual domains. This observation included the pSTS ROIs previously associated with audiovisual integration. Moreover, this effect remained even when activation in response to sensori-motor processing was removed. Therefore previous findings of enhanced activation in pSTS for crossmodal relative to uni-modal stimuli were not replicated. Second, although there was a main effect of congruency in bilateral caudate (congruent > incongruent sequential stimuli) there was no effect of congruency in the pSTS region of interest. In short, there was no evidence that pSTS activation was higher for audiovisual than uni-modal object matching. Instead, the results suggest that the pSTS is involved in the process of actively associating two objects both within and across modalities. Below, it is discussed how and why these findings differ from those previously reported.

4.51 Bimodal versus Uni-modal object matching

No region, including the pSTS ROIs, showed increased activation for bimodal relative to uni-modal stimuli when the task, attention and number of stimuli per trial were controlled. This conflicts with several previous studies that report pSTS activation is higher for bimodal than uni-modal inputs (Beauchamp *et al.,* 2004a, 2004b; Calvert *et al.,* 2000; van Atteveldt *et al.*, 2004; Wright *et al.,* 2003). It is suggested that these previous studies did not control for stimulus and attentional confounds. For example, Beauchamp *et al.* (2004b), using conceptual stimuli and tasks, compared activation for a series of audiovisual trials (i.e. two object stimuli per trial) to activation from a series of single uni-modal trials (i.e. one object stimulus per trial). Although the same objects were presented in each condition, the number of objects in audiovisual trials was double that for the uni-modal trials. The increased number of objects per trial has well recognised consequences on the haemodynamic response (Binder, *et al.,* 1994; Fox, 1989; Price *et al.*, 1992, 1996) which are sufficient to explain why audiovisual activation is higher than uni-modal activation when the number of stimuli per trial is not controlled.

4.52 Effect of Task

The association of pSTS with multi-modal integration could also result from task confounds during both active (Beauchamp *et al.,* 2004b) and passive tasks (Calvert *et al.,* 2000, 2001; van Atteveldt *et al.,* 2004; Wright *et al.,* 2003). For example, if pSTS is involved in making conceptual associations between incoming stimuli, subjects are more likely to make associations between two stimuli that arrive in close temporal proximity (as in the audiovisual conditions) than when single stimuli are separated in time. To avoid these confounds, a matching task on two sequentially presented stimuli was used in all conditions, thereby necessitating the comparison of two incoming stimuli. In this context activation in the pSTS ROI was the same for crossmodal and intra-modal matching. These findings are therefore more consistent with the designation of pSTS as an area specialised for amodal conceptual associations than an area that actively integrates auditory and visual inputs.

4.53 Congruent versus Incongruent trials

The results did not replicate previous reports that activation in the pSTS increases when audiovisual stimuli are congruent relative to incongruent (Calvert *et al.*, 2000). There are several points to note about the inter-study discrepancy reported here. First, in Calvert *et al.* (2000), the congruent and incongruent conditions were tested in separate experiments, which is likely to enhance attentional and strategic differences between conditions. It is therefore possible that, during incongruent audiovisual conditions, subjects selectively attended to one modality while actively suppressing attention to the other modality because, in the context of conflicting information, this strategy improved their ability to comprehend speech. As a consequence of this condition-specific strategy, activation would be lower for incongruent than congruent bimodal trials. Indeed, attention to one modality only during bimodal presentation has been found to illicit sub-additive effects (Talsma *et al.*, 2007; Talsma and Woldorff, 2005). In the study reported in this Chapter, subjects were engaged in a task that required them to attend to both the auditory and visual modalities. In this context, activation in the pSTS ROIs was unaffected by congruency or presentation modality.

4.54 Patchy organisation within human STS

In an elegant, high resolution fMRI study, Beauchamp and colleagues (Beauchamp *et al.*, 2004a) suggest that auditory and visual inputs arrive in STS in separate patches of cortex and are integrated in intervening cortex. This conclusion was based on observations that different patches of STS responded maximally to auditory and visual stimuli with intervening patches showing enhanced response to bimodal audiovisual stimuli than either auditory or visual stimuli alone. In the study reported in this Chapter, the voxels were 3x3x3mm (as opposed to 1.6x1.6x1.6mm in Beauchamp *et al.*, 2004a) which may have prevented the detection of "patchy" pSTS cortex. However, the results still call into question conclusions that these patches of amodal cortex actively integrate visual and auditory inputs. If these patches have enhanced responses to bimodal audiovisual stimuli then this should be detected even when the voxel size is larger (because more patches are responding to bimodal than uni-modal stimuli). Indeed, Beauchamp *et al.* (2004a) identified the region of interest for their high resolution study on the basis of multi-sensory activation in a study with low resolution (voxel size 3.75 x 3.75 x 5mm). The present finding that intra-modal auditory and visual matching activated these pSTS ROIs as much as crossmodal audiovisual matching therefore necessitates further investigation of patchy STS cortex using high resolution fMRI. Specifically, the effect needs to be replicated when attentional factors and the number of stimuli per trial are controlled. Only then can we exclude the possibility that enhanced activation for bimodal relative to uni-modal stimuli in patchy STS results from stimulus or attentional confounds.

4.6 Summary

In conclusion, this experiment differed from previous audiovisual studies because it attempted to control for both task and stimulus presentation parameters. Specifically, in both crossmodal and intra-modal conditions, subjects were instructed to compare two perceptually different meaningful stimuli. Under this context, equivalent pSTS activation was found for crossmodal and intra-modal stimulus matching. However, the inclusion of intra-modal stimuli in this experiment meant that the stimuli within each pair had to be presented sequentially. Simultaneous presentation was not used because when two auditory stimuli are presented at the same time, they interfere with one another at both perceptual and attentional levels (Jancke and Shah, 2002; Lipschutz *et al.,* 2002). It is therefore important to determine whether input synchrony is a determining factor in the involvement of pSTS in crossmodal integration. For example, although the effect of congruency may depend on the task and stimuli used, the nature of this dependency requires further investigation. For example, it may depend on whether the stimuli involve continuous speech or static objects (Calvert and Lewis, 2004). Moreover, previous studies have used either verbal stimuli *or* nonverbal stimuli, whereas in this experiment verbal and nonverbal information were always present in each trial. The following two chapters are designed to address two questions. Chapter 5 investigates differences in the activation pattern for simultaneous and sequential, as well as congruent and incongruent audiovisual pairs. In Chapter 6 the question is asked whether there are differences between the integration of verbal versus non-verbal stimuli, using the same crossmodal simultaneously presented conceptual stimuli.

Chapter 5

5 The influence of congruency and presentation rate on audiovisual object processing

5.1 Introduction

The previous chapter introduced the issues of congruency in audiovisual integration. To briefly summarise, previous studies reporting super-additive effects are confounded by differential stimulus intensity between uni-modal and crossmodal conditions, and the reporting of congruency effects are inconsistent and confounded by attentional modulation. When these confounds were controlled by engaging subjects in a task that required them to attend to both the auditory and visual modalities, activation in the pSTS ROIs was unaffected by congruency (i.e. no difference between matching congruent versus incongruent pairs) or presentation modality (i.e. no difference between audiovisual or intra-modal matching). However, it is important to determine whether input synchrony is a modulating factor for an effect of congruency.

In early studies, Calvert *et al.* (1997, 1999) observed that combined auditory and visual inputs increased activation in sensory specific areas but did not affect polysensory areas (e.g. left pSTS). These early studies used the numbers 1 to 10 that were either heard (auditory input), lip read (visual input) or both heard and lip read simultaneously. In contrast, in their later study that reported the super-additive congruent response in pSTS (Calvert *et al.*, 2000), the stimuli were speech streams which provided a longer temporal span for audiovisual integration. The implication from that study is that pSTS activation may depend on whether audiovisual stimuli are simultaneously or sequentially presented, as well as the time scale over which the synchrony takes place.

As indicated in the context of congruency highlighted in Chapter 4, differences between synchronised and desynchronised stimuli may depend on attention. When audiovisual stimuli are synchronised, then attention to both inputs enhances speech comprehension, but when audiovisual stimuli are desynchronised subjects may suppress the conflicting visual information thereby reducing the total stimulus processing and activation. The effect of audiovisual synchrony in pSTS has been investigated in several fMRI studies (Macaluso *et al.*, 2004; Miller and D'Esposito, 2005; Olson *et al.*, 2002; van Atteveldt *et al.*, 2007a). A review of the findings from these studies was presented in Table 4.1a, Chapter 4. To recap, no difference in pSTS activation was observed for synchronous and desynchronised 1) audiovisual speech streams (Olson *et al.*, 2002), 2) Faces and speech sounds (Miller and D'Esposito, 2005), or 3) letters and speech sounds (van Atteveldt *et al.*, 2007a). Macaluso *et al.* (2004) reported that simultaneous but not sequential audiovisual input increased left STS activation relative to the uni-modal conditions (spoken words and lip-reading) however this effect was ventral to the pSTS region of interest (see Table 4.1a, Chapter 4). Taken together, these findings are in conflict with those of Calvert *et al.* (2000).

In summary, although the effect of audiovisual synchrony and congruency on pSTS activation varies across studies, there has been a consistent consensus that pSTS is important for audiovisual integration of conceptual and phonetic stimuli even when static (temporally brief) stimuli are used. For example, pSTS activation has been observed for visually presented letters with their sounds (van Atteveldt *et al.*, 2004; 2007a), faces and speech sounds (Calvert *et al.*, 2000; Kreifelts *et al.*, 2007; Sekiyama *et al.*, 2003; Wright *et al.*, 2003), environmental sounds presented with static black and white photographs or line drawings (Beauchamp *et al.*, 2004b) and during a phoneme detection task on static speech relative to a closed mouth condition (Calvert and Campell, 2003).

The study reported in the preceding Chapter (hereafter referred to as Experiment 1) investigated the response characteristics of the pSTS when level of stimulus input and task demands were controlled. The findings suggested that pSTS activation reflected amodal processing (i.e. independent of stimulus modality), with no modulation by the congruency of the pair to be matched. It was concluded in that Chapter that previous reports of increased pSTS activation in crossmodal processing was due to the confounding effect of attention between congruent and incongruent pairs: i.e. subjects attended to only one stimulus input when there was a conflict between crossmodal inputs. The question asked in this Chapter then, is what the differences in response characteristics of pSTS are when attention to both crossmodal inputs is controlled but presentation onset asynchrony is manipulated.

This study

The inclusion of uni-modal stimuli in Experiment 1 meant that the individual stimuli within a pair had to be presented sequentially (one after the other). Simultaneous presentation was not used because when two auditory stimuli are presented at the same time, they interfere with one another at both perceptual and attentional levels

(Jancke and Shah, 2002; Lipschutz *et al.*, 2002). The experiment reported in this Chapter (hereafter referred to as Experiment 2) was conducted using the same audiovisual stimuli as Experiment 1 but with simultaneous rather than sequential presentation. By combining the crossmodal data from these two experiments, any differences in pSTS response pattern between audiovisual pairs that are congruent or incongruent when presented simultaneously or sequentially can be investigated.

5.2 Method

Subjects

There were 18 subjects in Experiment 1 (11 women, 7 men, mean age 26), and 18 subjects in Experiment 2 (12 women, 6 men, mean age 26). All were right handed native English speakers with normal or corrected to normal vision. All had normal neurological and audiological status. The study was approved by the joint ethics committee of the Institute of Neurology and University College London Hospital, London, UK.

5.21 Materials and Methods

Experiment 1: Sequential presentation

Full details on the stimulus, design and acquisition parameters for the audiovisual conceptual matching experiment can be found in Chapter 4. To briefly recap, the experiment used four types of object stimuli: pictures of objects, their written names, their auditory names and their associated environmental sounds (i.e. two visual and two auditory). This resulted in four different conditions, with an equal number of congruent and incongruent trials within each condition:

(1) Crossmodal trials with pictures of objects and spoken names,

(2) Cross modal trials with written names and environmental sounds,

(3) Intra-modal visual trials with one picture and one written name, and

(4) Intra-modal auditory trials with one spoken object name and one environmental sound.

Subjects made a match or no match decision on each trial using a key press response. For the experiment reported here, the contrast of interest was the difference between activation for crossmodal congruent and incongruent conditions (conditions 1 and 2 combined).

Experiment 2: Simultaneous presentation of bimodal objects

In Experiment 2, subjects were presented bimodally with two simultaneously presented object stimuli, one in the visual modality (colour photograph or written object name) and one in the auditory modality (spoken object name or environmental sounds). As in Experiment 1, they were instructed to indicate, via a key pad response, whether the two stimuli referred to the same object or not. The 108 object stimuli were identical to those used in Experiment 1 (see Appendix 2 for a complete list). In total there were four different crossmodal combinations:

(1) One photograph with one spoken name,

(2) One written name with one environmental sound,

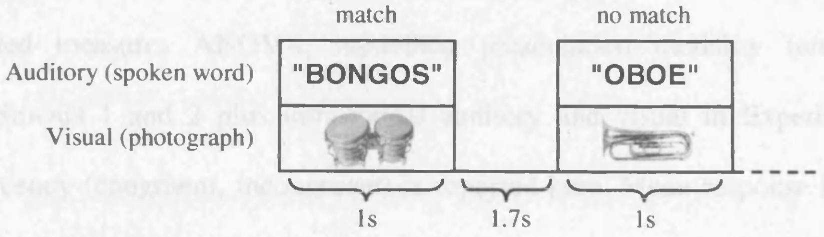
(3) One written name with one spoken name (i.e. verbal only stimuli),

(4) One photograph with one environmental sound (i.e. nonverbal only stimuli).

As in Experiment 1, the stimulus conditions were blocked with 3 congruent and 3 incongruent trials per block. Trial length was 2.7 (1s stimulus duration followed by 1.7s fixation), block length was 16.2s (2.7 x 3 congruent and 3 incongruent trials), fixation after each block alternated between 1.62s and 16.2s and there were a total of 24 blocks in each of 4 different scanning sessions. Although the present comparison between experiments involves the 2 stimulus combinations that corresponded to those in Experiment 1 (photograph-spoken word, written word-environmental sound), Experiment 2 also included 2 other audiovisual matching conditions (photograph-environmental sound, written word-spoken word). These are reported in the following chapter (Chapter 6). See Figure 5.1 for an example of the simultaneous stimulus trials relevant to the analysis reported here.

Figure 5.1: Stimulus trials for simultaneous presentation in Experiment 2

Audiovisual spoken word – picture matching



Audiovisual environmental sound – written word matching

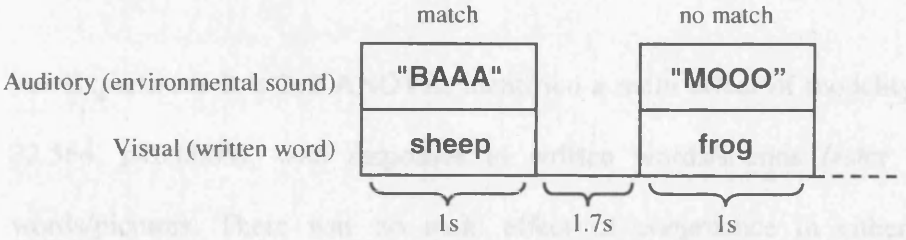


Figure provides an example of congruent and incongruent trials for simultaneous presentation of auditory and visual inputs. For the comparison between simultaneous and sequential audiovisual matching, only those simultaneous trials using a combination of [spoken word-picture] and [environmental sound–written word] are used in the current analysis.

To investigate any differences between sequential and simultaneous matching, the present analysis was based on the effect of congruent vs. incongruent in (a) the two cross modal (AV) conditions in Experiment 1, and (b) the corresponding crossmodal conditions in Experiment 2 (spoken word + photograph; environmental sound + written word). This allowed testing of the main effect of congruency (congruent vs. incongruent) and its interaction with temporal presentation (simultaneous vs. sequential). In addition, a comparison was carried out between the Z-scores for all peak clusters for simultaneous compared with sequential matching.

5.3 Behavioural Results

Reaction time data for Experiment 1 was reported in detail in Chapter 4. A separate analysis on the reaction times for all four conditions presented in Experiment 2 is given in the following Chapter (6). For the purposes of the current analysis, a repeated measures ANOVA, modelling presentation modality (crossmodal in Experiments 1 and 2 plus intra-modal auditory and visual in Experiment 1) and congruency (congruent, incongruent) is reported here. Mean response latencies and standard deviations for the relevant contrasts for this Chapter are shown in Table 5.1.

For Experiment 2, a 2x2 ANOVA, identified a main effect of modality (F[1,17]) = 32.564, $p<0.0005$), with responses to written words/sounds faster than spoken words/pictures. There was no main effect of congruence in either experiment (Expt.1: $p=0.423$; Expt.2: $p=0.806$), and no interaction between presentation modality and congruency (Expt.1: $p=0.562$; Expt.2: $p=0.534$).

Table 5.1: Behavioural data

Experiment	*Matching Condition*	*Mean (ms)*	*SD*
1. Sequential:	AV con	1001	214
	AV inc	1029	211
	VV con	927	246
	VV inc	927	250
	AA con	1114	166
	AA inc	1122	168
2. Simultaneous:	AsVw con	945	191
	AsVw inc	946	176
	AwVp con	875	153
	AwVp inc	870	146

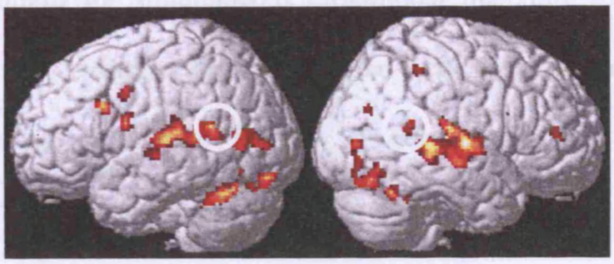
Mean and standard deviation for reaction times in response to audiovisual, visual and auditory matching tasks in Experiments 1a and 2. Data are for 10 subjects in Experiment 1 (due to technical difficulties with recording from the keypad) and for 18 subjects in Experiment 2. Sequential = Experiment 1, Simultaneous = Experiment 2, con = congruent trials, inc = incongruent trials, AV = audiovisual matching, VV = visual matching, AA = auditory matching, Aw= Auditory words, As = Auditory sounds, Vw = Visual words, Vp= Visual pictures.

5.4 Functional Imaging Results

5.41 The effect of congruency during simultaneous and sequential crossmodal matching

Across experiments, there were no significant effects of congruent > incongruent. However, incongruent > congruent activated a distributed set of bilateral regions that included the left and right pSTS ROIs, see Table 5.2 and Figure 5.2. The peak pSTS effects lay lateral and slightly superior [-60, -52, 14/+64, -48, 12] to the defined regions of interest [+/-50, -52, 8] and were primarily driven by Experiment 2 (simultaneous presentation) with no significant effect of incongruency in Experiment 1 (sequential presentation), see Figure 5.3 and Table 5.2.

Figure 5.2: Incongruent > congruent matching of simultaneously presented audiovisual objects



Activation rendered on an averaged surface model of the brain with an uncorrected threshold of p<0.001 and a minimum cluster size of 20 voxels.

Table 5.2: Incongruent relative to congruent trials for simultaneous matching

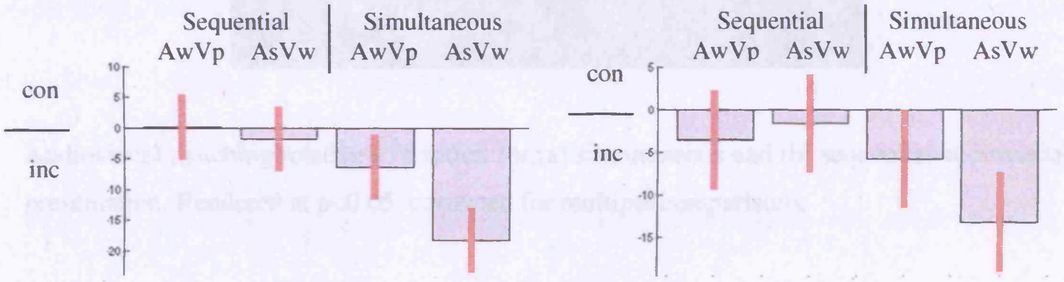
Anatomical region	*Co-ordinates*			*Z scores:*		
	Simultaneous			*Simul.*	*Sequl.*	*Interaction*
L superior temporal gyrus/sulcus	***-60**	**-52**	**14**	**5.0**	**ns**	**3.5**
	-56	-22	8	5.1	ns	4.0
	-62	-28	12	5.7	ns	4.1
	-62	-42	10	5.1	1.7	2.9
R superior temporal gyrus/sulcus	***64**	**-48**	**12**	**3.6**	**ns**	**ns**
	64	-12	2	6.6	ns	5.3
	50	-14	4	5.1	ns	3.3
	62	-16	-8	5.1	ns	3.8
	40	-18	-8	5.4	ns	4.0
	46	-24	16	5.2	ns	3.3
	46	-30	-2	5.7	ns	3.9
	56	-34	4	5.3	ns	3.4
R occipital	38	-80	4	5.4	ns	4.4
	38	-70	-18	5.3	ns	3.2
L occipital	-32	-84	-16	5.2	ns	3.8
	-50	-80	6	5.1	ns	3.8
L cerebellum	-38	-60	-24	5.3	ns	4.0
	-34	-58	-22	5.1	ns	3.5

Peak co-ordinates of significant clusters for matching incongruent relative to congruent trials with simultaneous (p<0.05 corrected for multiple comparisons across the whole brain) and sequential (p<0.05 uncorrected) audiovisual input and the interaction between congruence and synchrony. Co-ordinates highlighted in bold and marked with an asterisk are those closest to our left and right pSTS ROI. Key: L = left, R = right, ns = not significant.

5.42 The effect of simultaneous versus sequential crossmodal matching

The activation pattern for simultaneous crossmodal matching relative to fixation (Experiment 2) was virtually identical to the activation pattern for sequential crossmodal matching relative to fixation (Experiment 1), see Figure 5.4 and Appendix 4 for co-ordinates and corresponding Z-scores of all peak clusters.

Figure 5.3: Effect sizes in bilateral pSTS for all conditions in each Experiment

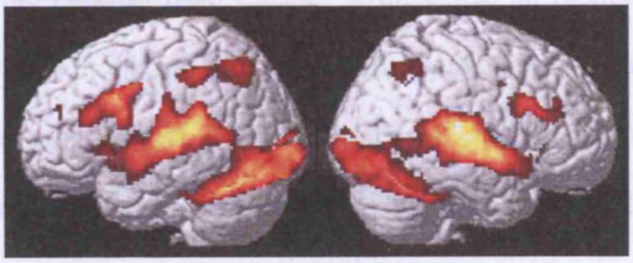


a. Left pSTS **b. Right pSTS**

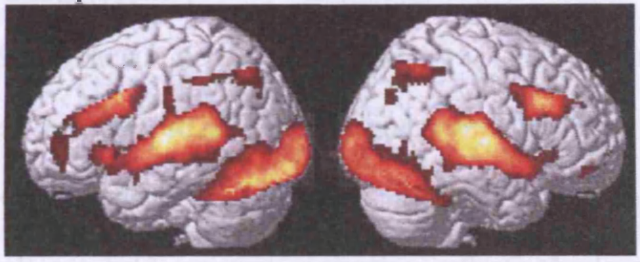
Effect sizes of activation at voxels in the left [-60, -52, 14] and right [64, -48, 12] pSTS, showing the effect of congruency in Experiment 2 (simultaneous) compared with Experiment 1 (sequential). NB positive scale on *y*-axis indicates effect size for congruent matching, negative *y*-axis for incongruent matching.

Figure 5.4: Comparison of audiovisual matching in Experiments 1 and 2

a. Simultaneous



b. Sequential



Audiovisual matching relative to fixation for (a) simultaneous and (b) sequential audiovisual presentation. Rendered at $p<0.05$, corrected for multiple comparisons.

5.5 Discussion

Contrary to previous reports of increased activation for congruent audiovisual stimuli, Experiment 2 showed increased activation for incongruent relative to congruent audiovisual inputs throughout a widely distributed network of regions that included the pSTS. Second, it was shown that the network of brain regions activated for sequential audiovisual matching in Experiment 1 included all the areas that were activated by matching simultaneously presented audiovisual stimuli in Experiment 2. Below these differences are discussed in the context of previous reports for both congruency and synchrony in audiovisual integration.

5.51 Congruent versus incongruent matching

Consistent with the findings from Chapter 4, there was no replication of previous reports that activation in the pSTS increases when audiovisual stimuli are congruent relative to incongruent (Calvert *et al.,* 2000). To the contrary, these results were in the opposite direction, with a network of regions, including pSTS, activated more strongly for incongruent than congruent pairs. In these two experiments, subjects were engaged in a task that required them to attend to both the auditory and visual modalities. In this context, increased activation was observed in a network of regions, including the pSTS ROIs, for incongruent relative to the congruent trials when stimuli were simultaneously presented. This suggests that, even when the number of stimuli are held constant, activation reflects the level of conceptual processing, which is greater when two incongruent objects are simultaneously presented (two concepts) than when two congruent objects are simultaneously presented (one concept). Consistent with this hypothesis, activation in the pSTS ROIs was unaffected by congruency or presentation modality in Experiment 1 where object concepts were presented one at a time in each condition.

In summary, previous studies have either shown a positive effect of congruency in pSTS (greater activation for congruent than incongruent; Calvert *et al.,* 2000) or no effect of congruency (e.g. Hein *et al.*, 2007; Ojanen *et al.,* 2005; Taylor *et al.*, 2006; van Atteveldt *et al.,* 2007a, 2007b). These data have added to this inconsistency by demonstrating a negative effect of congruency (greater activation for incongruent than congruent) when two stimuli are simultaneously presented. The most likely interpretation is that the effect of congruency depends on the task and stimuli used. However, the nature of this dependency requires further investigation. For example, it may depend on whether the stimuli involve continuous speech or static objects (Calvert and Lewis, 2004). Alternatively, it could be hypothesised that if subjects are able to attend to one input modality while suppressing the other, then activation will be less for incongruent bimodal trials. In contrast, if subjects are forced to attend to both modalities then activation will be higher for incongruent bimodal trials that carry twice the conceptual and phonetic information content as congruent trials.

5.52 Simultaneous versus Sequential Presentation

It could also be argued that because moving visual stimuli were not presented with the simultaneously presented spoken words and environmental sounds, and hence provide a truly synchronous event, I was unable to detect enhanced pSTS activation for bimodal relative to uni-modal stimuli. However, as reviewed in the introduction, the association of the pSTS ROIs with multi-modal integration is not limited to moving stimuli. Moreover, the stimulus and task factors highlighted in the context of non-moving conceptual stimuli also apply to results from studies that did use moving speech stimuli. In short, these results are based on temporally brief verbal and nonverbal conceptual stimuli that could only be integrated at a late level of processing because there was no correspondence at a perceptual level. Nevertheless, they call into question previous conclusions based on both moving and non-moving stimuli. Further studies are therefore required to determine if the pSTS ROIs are activated by bimodal more than uni-modal processing of continuous and synchronous audiovisual speech streams when attention and stimulus input are controlled. Such an experiment might involve the comparison of (1) synchronous versus asynchronous audiovisual speech when attention is controlled (e.g. if subjects were instructed to press a button when there was a mismatch in the auditory and visual inputs); and (2) audiovisual speech versus asynchronous intra-modal matching (e.g. deciding if mouth movements correspond to written text).

5.6 Summary

The same bilateral network observed for sequential matching in Experiment 1 was activated in Experiment 2 when the equivalent audiovisual stimuli were simultaneously presented. However, in this context, greater activation was observed in the pSTS for incongruent relative to congruent crossmodal matching condition. Activation in the pSTS ROI may therefore depend on the number of objects that are simultaneously attended to (two different objects during incongruent audiovisual trials versus one object during congruent audiovisual trials). This would explain why the effect of congruency in the pSTS ROI can reverse in other tasks that do not control attention across modalities (van Atteveldt *et al.*, 2004). Taking all these results into account, the conclusion is that pSTS activation is involved in the amodal conceptual association of two objects. This process may play an important role in audiovisual integration although I found no evidence that this region is more activated by audiovisual than uni-modal conceptual processing.

Chapter 6

6 Verbal versus nonverbal audiovisual object processing

6.1 Introduction

In the previous two chapters the focus of attention was placed on the role of the posterior STS in crossmodal audiovisual integration. This is an area that has been associated with audiovisual processing in experiments that used either verbal (Calvert *et al.,* 2000; Kreifelts *et al.,* 2007; Macaluso *et al.,* 2004; Raij *et al.,* 2000; Sekiyama *et al.,* 2003; van Atteveldt *et al.,* 2004, 2007a; Wright *et al.,* 2003) or nonverbal (Beauchamp *et al.,* 2004b; Taylor *et al.,* 2006) stimuli. The current chapter focuses on how audiovisual integration differs for verbal and nonverbal stimuli. The term verbal refers to word processing – whether written, spoken or lip-read. Word stimuli can either be presented in the form of continuous speech (as in Calvert *et al.,* 2000; Macaluso *et al.,* 2004) or in the form of single words (Ojanen *et al.,* 2005; Olson *et al.,* 2002; Raij *et al.,* 2000; van Atteveldt *et al.,* 2004, 2007a). The verbal stimuli used in this chapter were written and spoken object names because this permitted a controlled comparison to nonverbal audiovisual stimuli in the form of pictures of objects and the environmental sounds associated with them. Purely verbal audiovisual stimuli (spoken and written words) could therefore be compared to purely nonverbal stimuli (pictures and environmental sounds) while controlling for stimulus properties by including control conditions that involved one verbal and one nonverbal stimulus (spoken words with pictures or written words with environmental sounds).

In both the neuroimaging and neuropsychological literatures, verbal and nonverbal inputs are generally referred to as being from different modalities, often within the visual perceptual modality. However, the term "modality" is also used to refer to the type of stimulus (e.g. auditory versus visual). Since the verbal versus nonverbal distinction is not specific to auditory and visual input, the term "modality" will only be used to refer to the type of perceptual input, that is, whether the sensory input is in the visual or auditory domain.

Verbal and nonverbal stimuli *both* access phonological and semantic processes. However, they do so in different ways. For verbal stimuli, phonetic analysis of the input precedes recognition at the semantic level (e.g. Indefrey and Levelt, 2000, 2004). By contrast, for nonverbal stimuli, semantic processing precedes phonological retrieval (e.g. Glaser and Glaser, 1989; Seifert, 1997). Consequently, in an audiovisual matching task, purely verbal stimuli (e.g. auditory and visual words) can be equated at a phonological level prior to accessing semantics whereas purely nonverbal stimuli (e.g. pictures and environmental sounds) can be equated at a semantic level prior to accessing phonology. Audiovisual matching of verbal and nonverbal stimuli may therefore differentially engage phonological and semantic processing regions respectively. In addition, dissociations in verbal and nonverbal semantic processing (see below for a review) may also influence activation associated with verbal and nonverbal audiovisual integration. Below, I briefly review the relevant literature on verbal and nonverbal semantic processing and the influence of these findings on my anatomical hypotheses.

6.11 Dissociating verbal and nonverbal processing

A dissociation between verbal and nonverbal semantic processing has been suggested by reports of brain damaged patients with deficits selective for one or other stimulus type. Depending on the theoretical perspective taken, this has been cited as evidence for separate visual and verbal semantic systems (e.g. Ferreira *et al.*, 1997; Warrington, 1975; Warrington and McCarthy, 1994), or for a shared distributed semantic system differentiated by the type of knowledge primarily involved during acquisition (e.g. Saffran *et al.*, 2003). In contrast, rather than interpreting deficits in terms of stored knowledge, differences in verbal or nonverbal abilities have been associated with lateralisation of the *processes* engaged during access to stored knowledge, with the left hemisphere more often engaged in accessing verbal information and the right hemisphere more engaged in accessing nonverbal information (Coslett and Saffran, 1992; for reviews see Caramazza *et al.*, 1990 and Lambon-Ralph *et al.*, 1999). Unfortunately, most of the evidence from patient data comes from a comparison of verbal/nonverbal processing in the visual modality only. Therefore, the conclusions concerning amodal or separable verbal/nonverbal systems are limited.

Recently, functional neuroimaging of normal subjects has provided another source of evidence for a dissociation between verbal and nonverbal processing within either the auditory modality or the visual modality (Adams and Janata, 2002; Bright *et al.*, 2004; Chee *et al.*, 2000; Dick *et al.*, 2007; Giraud and Price, 2001; Humphries *et al.*, 2001; Perani *et al.*, 1999; Thierry *et al.*, 2003; Thierry and Price, 2006; Vandenberghe *et al.*, 1996; von Kriegstein *et al.*, 2003). Critically, however, the areas associated with verbal and nonverbal stimuli differ according to the input modality (visual or auditory). These within-modality effects are difficult to interpret because they are confounded by perceptual differences in the nature of the verbal and nonverbal stimuli. To circumvent perceptual confounds, Thierry and Price (2006) looked for verbal versus nonverbal processing differences that were independent of stimulus modality. Combining data from one experiment using auditory stimuli and another using corresponding visual stimuli, they reported a left/right double dissociation for verbal/nonverbal material, independent of sensory modality. Specifically, verbal relative to nonverbal material activated anterior and posterior regions of the left STS and the ventral left inferior frontal gyrus while nonverbal relative to verbal material activated the right mid fusiform and pMTG. Of relevance here is that the right mid-fusiform gyrus activated in Thierry and Price (2006) for nonverbal relative to verbal semantic processing was also activated in Chapter 3 of this thesis for increased perceptual input (form and colour and sound) during nonverbal object naming (pictures or environmental sounds). Therefore, I have already presented some evidence that the right mid-fusiform might be involved in nonverbal audiovisual integration.

The anatomical dissociation reported in Thierry and Price (2006) provides hypotheses for the current experiment. However, it should still be noted that the functional level at which the verbal versus nonverbal differences arise in Thierry and Price (2006) is debatable. For example, the right posterior superior temporal region associated with nonverbal conceptual processing in Thierry and Price (2006) has been associated with spatial localisation in both the auditory (Rauschecker, 1998a; 1998b; Rauschecker and Tian, 2000) and visual (Milner and Goodale, 1993) domains. Conversely, auditory and visual verbal stimuli are more likely to evoke speech production and morpho-syntactic associations compared to nonverbal stimuli. Indeed, the left posterior superior temporal area that was activated for verbal relative to nonverbal conditions in Thierry and Price (2006) has previously been associated with speech production processes (Blank *et al.,* 2002; Hickok *et al.,* 2003; Okada *et al.,* 2003; Price *et al.,* 2006; Warren *et al.,* 2005; Wise *et al.,* 2001). Likewise, the left anterior superior temporal cortex that was activated for verbal relative to nonverbal conditions in Thierry and Price (2006) has previously been associated with morpho-syntactic processing (Bornkessel *et al.*, 2005; Dronkers, 2000; Dronkers *et al.,* 2004; Friederici and Kotz, 2003; Friederici *et al.,* 2003; Humphries *et al.*, 2005; Stowe *et al.,* 1999; Vandenberghe *et al.*, 2002)

6.12 This study

The present study contrasts the effects of matching verbal versus nonverbal simultaneously presented audiovisual pairs by manipulating the type of material. Verbal stimuli included auditory words (Aw) and visual words (Vw); nonverbal stimuli included environmental sounds (As) and pictures of objects (Vp) (where "A" indicates auditory presentation, "V" indicates visual presentation, "s" indicates sounds and "p" indicates pictures). It is predicted that within the shared semantic network engaged by audiovisual stimuli (clearly illustrated in Chapter 4), a dissociation will be observed between these different types of stimulus material. Previous studies have associated phonological processing with the left superior temporal sulcus (Binder *et al.,* 2000; Scott *et al.,* 2000; Wise *et al.,* 2001), the left supramarginal gyrus and left posterior inferior frontal gyrus (Devlin *et al.,* 2003; Mummery *et al.*, 1999a; Paulesu *et al.*, 1993; Price *et al.,* 1999). In contrast, semantic processing has been associated with the left middle temporal gyrus (Binder *et al.*, 1997), left anterior temporal lobe (Scott *et al.*, 2000, Vandenberghe *et al.*, 1996) and the angular gyri (Mummery *et al.*, 1999a).

6.2 Method

6.21 Materials and Method

Subjects

18 subjects participated in this Experiment (12 women, 6 men, mean age 26). All were right handed native English speakers with normal or corrected to normal vision and gave informed consent to take part. All had normal neurological and audiological status. The study was approved by the joint ethics committee of the Institute of Neurology and University College London Hospital, London, UK.

Experimental design and stimuli

Details are as provided in the chapter 5, and the experimental paradigm is shown in Figure 6.1. To briefly recap, subjects were presented with two simultaneously presented meaningful stimuli, one in the visual modality (colour photograph or written object name) and one in the auditory modality (spoken object name or environmental sounds). They were instructed to indicate, via a key pad response, whether the two stimuli referred to the same object or not. These audiovisual pairings resulted in four different conditions, with differing levels of verbal components:

(1) Cross modal trials with written names and spoken names (i.e. two verbal components: AwVw),

(2) Crossmodal trials with pictures of objects and environmental sounds (i.e. no verbal components: AsVp),

(3) Crossmodal trials with pictures of objects and spoken names, and (i.e. one verbal component: AwVp),

(4) Cross modal trials with written names and environmental sounds (i.e. one verbal component: AsVw).

The effects of interest were 1) the number of verbal components in crossmodal integration, and 2) the influence of congruency on these pairings.

Figure 6.1: Stimulus trials for simultaneous audiovisual matching

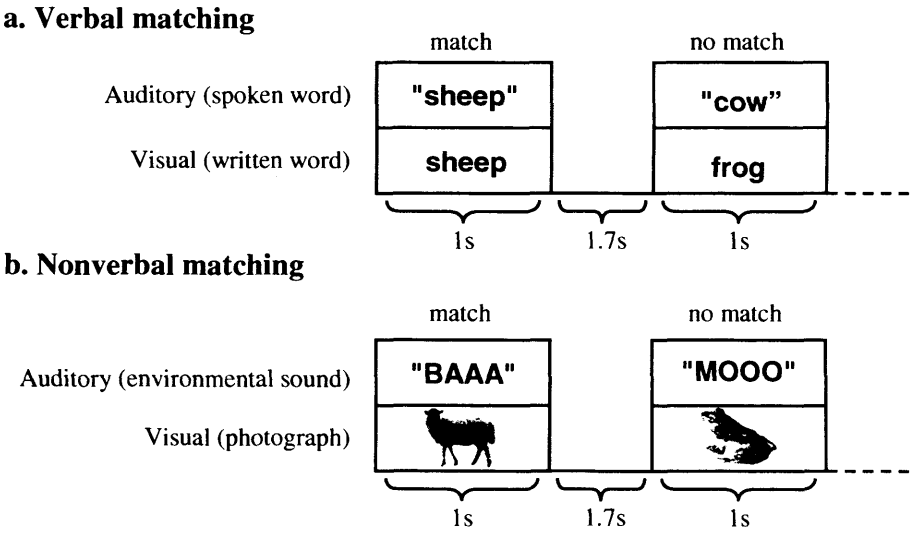


Figure 6.1a: stimulus trials with maximum verbal material, and 6.1b: stimulus trials with no verbal material. Subjects made a "yes they match" or "no they don't match" response using a keypad depending on whether auditory and visual stimuli referred to the same object or not.

6.22 Data Acquisition

See Chapter 5 for acquisition parameters.

Data Analysis

Functional data were analysed with statistical parametric mapping (SPM2, Wellcome Trust Centre for Neuroimaging, London, UK) as reported previously.

First level statistical analyses (single subject and fixed effects) modelled each trial type independently by convolving the onset times with the haemodynamic response function. There were 8 trial types: 4 conditions x 2 congruency. These 8 parameter estimates were then fed into a second level ANOVA. The analysis enabled the investigation of:

1. The effect of matching audiovisual pairs with different levels of verbal components. This used a parametric contrast relative to the amount of verbal or non-verbal information in each condition. For example the AwVw condition contained two items of verbal input whereas AwVp and AsVw contained one piece of verbal information and AsVp contained no verbal input,
2. The effect of congruency (congruent versus incongruent) on audiovisual matching,
3. The interaction between congruency and verbal content.

6.3 Behavioural Results

Reaction times were analysed using a repeated measures ANOVA, modelling presentation modality and congruency. Means and standard deviation are shown in Table 6.1. The 4x2 ANOVA (4 types of matching, 2 congruency) identified a main effect of condition (F[1,17]) = 32.564, $p<0.0005$). Post hoc analysis of this effect revealed faster responses when the stimuli included 1) spoken words than environmental sounds (*t* = 12.999; p<0.0005) and 2) written words than pictures (*t* = 5.631; p<0.0005). As a consequence of this additive effect, response latencies were faster for purely verbal audiovisual pairs (AwVw) relative to all other conditions. There was no main effect of congruence, but there was an interaction between presentation modality and congruency (F[3,51] = 15.675, p<0.0005). This effect was driven by faster responses to congruent than incongruent pairs, but only for the AwVw condition (*t*=-5.986; p<0.0005) with no effect of congruency in any other condition (p>0.05).

Table 6.1: Behavioural data

Matching Condition	*Verbal input*	*Mean (ms)*	*SD*
AwVw con	2	776	166
AwVw inc	2	834	161
AsVw con	1	945	191
AsVw inc	1	946	176
AwVp con	1	875	153
AwVp inc	1	870	146
AsVp con	0	964	195
AsVp inc	0	950	169

Table gives means and standard deviation for reaction times in response to audiovisual matching tasks with different levels of verbal input for congruent and incongruent trials. Key: con = congruent trials, inc = incongruent trials, A = auditory, V = visual, w = word, p = picture, s = sound, SD = standard deviation, ms = milliseconds.

6.4 Functional Imaging Results

All conditions activated a bilateral network of occipital, temporal, parietal and frontal regions (See Figure 6.2). This network reflects all stages of the semantic task from early sensory audiovisual input, through semantic processing, decision-making in response to the task, motor preparation and response execution. Nothing reached a corrected level of significance for congruent versus incongruent pairs across the whole brain, or in any of the regions of interest, even for the AwVw condition where an effect of congruency was observed behaviourally. Thus the following results for the verbal versus nonverbal comparison sum over the effects of congruency.

Figure 6.2: Common network for audiovisual integration

Figure shows common effects for all conditions relative to the fixation baseline, rendered on the SPM standard surface model of an averaged brain at $p<0.05$ corrected for multiple comparisons.

6.41 The effect of verbal information on audiovisual matching

To identify the effect of verbal information on audiovisual pairs, a parametric contrast was used, with contrast weights corresponding to the number of verbal components in each condition (see Method section). This identified increased activation in bilateral superior temporal gyri/sulci (see Table 6.2). To constrain the analysis further, the same parametric contrast was then inclusively masked (at a threshold of $p<0.05$ uncorrected) with the conditions involving one verbal component relative to no verbal components [AwVp > AsVp; AsVw > AsVp]. One significant cluster remained in the left STS at [-62, -36, 4; Z=4.9] (see Figure 6.3).

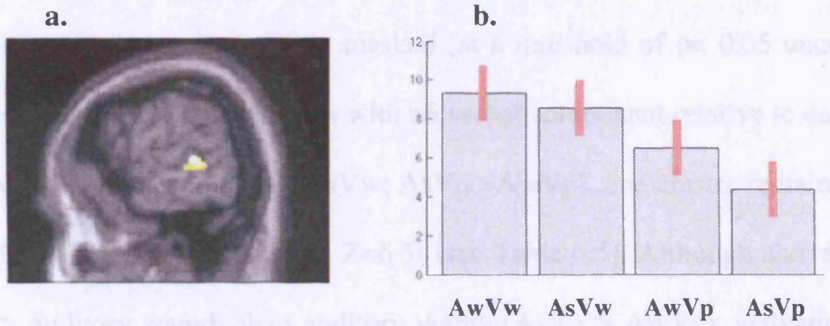
Table 6.3 shows that this region was more engaged by all conditions involving a verbal component relative to no verbal component (AsVp).

Table 6.2: Main effect of maximum verbal information

Anatomical region		x	y	z	Z-score
Left STS:	posterior	-62	-36	4	4.9
	middle	-64	-22	-2	4.0
	anterior	-58	-8	-6	3.4
Right STS:	posterior	64	-2	-6	4.7
	middle	66	-14	-2	4.6
	anterior	58	-24	-2	4.2

Table shows anatomical regions, MNI co-ordinates and corresponding Z-scores for significant clusters of activation for the main parametric effect of verbal stimuli.

Figure 6.3: Effect of verbal input on audiovisual matching



a. the effect of increasing verbal inputs rendered on an averaged T1-weighted section in the saggital plane (x=-62), with a threshold of p<0.001 uncorrected. Cluster centred at [-62, -36, 4].

b. Plot of parameter estimate at [-62, -36, 4] showing progressive increase in activation with verbal content. Effect sizes are the mean of the beta value summed over congruent and incongruent trials.

Table 6.3: Simple main effects for verbal information in posterior STS

AwVw > AsVp				AsVw > AsVp				AwVp > AsVp			
x	*y*	*z*		*x*	*y*	*z*		*x*	*y*	*z*	
-60	-38	2	(3.6)	-64	-40	0	(2.1)	-62	-36	2	(3.3)

Table gives anatomical co-ordinates (Z-scores in parentheses) showing level of activation in the contrast of each condition involving a verbal component, relative to matching with no verbal information. Co-ordinates given are peaks that fall within a sphere of 6mm centred on the main effect of verbal information [-62, -36, 4].

6.42 The effect of nonverbal input on audiovisual matching

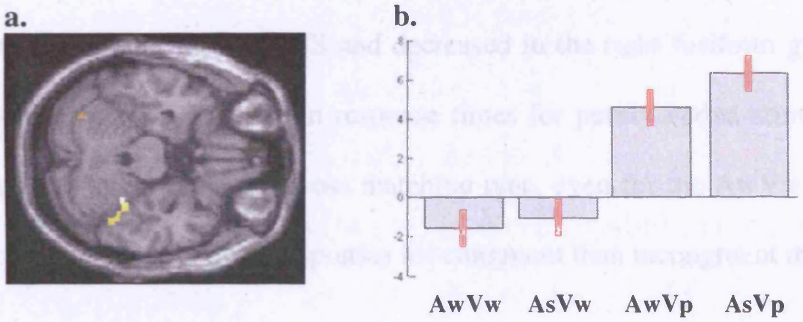
The effect of decreasing verbal content identified increased activation in bilateral occipito-temporal regions (see Table 6.4 for co-ordinates of significant clusters). When this contrast was inclusively masked (at a threshold of $p< 0.05$ uncorrected) with the simple effects of conditions with no verbal component relative to each of the others [AsVp > AwVw; AsVp > AsVw; AsVp >AwVp], one cluster remained in the right middle fusiform [30, -42, -24; Z=6.5] (see Table 6.5). Although activation was higher for auditory sounds than auditory words (AsVp > AwVp), activation in the right mid-fusiform area was primarily driven by the presence of pictures (see Figure 6.4).

Table 6.4: Main effect of nonverbal matching

Anatomical region	x	y	z	Z-score
Right fusiform gyrus	30	-42	-22	6.1
	30	-54	-16	5.7
	32	-62	-14	5.5
Left fusiform gyrus	-28	-60	-14	6.0
	-28	-78	-10	4.8
	-30	-42	-20	3.5
Right middle occipital gyrus	32	-94	10	5.7
	42	-90	4	3.2
Left middle occipital gyrus	-30	-94	18	5.3
	-36	-86	6	4.2
	-44	-84	4	4.0

The anatomical regions, MNI co-ordinates and corresponding Z-scores for activation that increased with decreasing verbal input.

Figure 6.4: Effect of no verbal input on audiovisual matching



a. Activation for matching pairs with no verbal components (AsVp), rendered on an averaged T1-weighted section in the axial plane (z= -24), with a threshold of p<0.001 uncorrected. Cluster centred at [30, -42, -24].

b. Plot of parameter estimate in right fusiform showing the effect is progressively driven by the loss of verbal information. Effect sizes are the mean of the beta value summed over congruent and incongruent trials.

Table 6.5: Nonverbal effects in the right fusiform

Nonverbal input				AsVp > AwVp				AsVp > AwVw				AsVp > AsVw			
x	y	z		x	y	z		x	y	z		x	y	z	
30	-42	-24	(6.5)	32	-42	-24	(1.8)	30	-44	-22	(7.0)	30	-42	-24	(7.3)

Table shows the effect of matching without verbal information (AsVp) relative to all other conditions, using a 6mm search volume based on the right fusiform peak. Co-ordinates and corresponding Z-scores (in parentheses) show that the effect is primarily driven by matching trials that involve pictures.

6.5 Discussion

The current study investigated the effect of verbal versus nonverbal material on activation during an audiovisual matching task. To summarise, within the bilateral network activated during audiovisual matching across all pair types, activation was modulated by the level of verbal content. When verbal material was maximal, activation increased in the left STS and decreased in the right fusiform gyrus. This was accompanied by a decrease in response times for purely verbal stimuli. There were no effects of congruency across matching type, even for the AwVw condition that showed significantly faster responses for congruent than incongruent trials.

6.51 Verbal audiovisual matching

My *a priori* prediction was that purely verbal stimuli could be matched at a phonological level whereas purely non-verbal stimuli could be matched at a semantic level. Therefore differential activation for verbal and nonverbal stimuli would be observed in phonological and semantic processing regions respectively. Remarkably, the peak co-ordinates for the effect of verbal versus nonverbal audiovisual matching

[-62, -36, 4] were identical to those reported by Thierry and Price (2006) for verbal relative to nonverbal semantic processing within either the visual or auditory domain [-62, -36, 4]. The same area has been associated with amodal phonological processing in a number of different studies (Blank *et al.,* 2002; Hickok *et al.,* 2003; Okada *et al.,* 2003; Price *et al.,* 2006; Warren *et al.,* 2005; Wise *et al.,* 2001). In other words, the results of this chapter confirm the *a priori* prediction that phonological processes play a greater role in verbal than nonverbal audiovisual matching.

6.52 Nonverbal audiovisual matching

Given the remarkable correspondence of the verbal effects reported here for audiovisual matching and in Thierry and Price (2006) for within modality semantic tasks, one might also predict that there would be a correspondence for the nonverbal effects. At first glance, this appeared to be the case. Nonverbal relative to verbal audiovisual matching increased activation in the right fusiform gyrus which was also reported for nonverbal relative to verbal stimuli during within modality tasks (Thierry and Price, 2006). Closer inspection of the activations however, reveals that the anatomical location of the two results is not the same. During audiovisual matching the nonverbal effect was centred at [30, -42, -24] on the medial surface of the fusiform gyrus but in Thierry and Price, the nonverbal effect is at [46, -46, -22] on the lateral surface of the fusiform gyrus. Likewise, the effect of combined audiovisual input reported in Chapter 3 was located laterally at [42, -40, -14].

Why might the right medial fusiform gyrus be more activated by audiovisual matching of nonverbal stimuli? This region is not classically associated with semantic processing. However, it has been shown to express a differential sensitivity to object category, with increased activation for pictures of artefacts than animals (Chao *et al.,* 1999, 2002). A recent study by Noppeney *et al.,* (2006) demonstrated that the category effects in the medial fusiform gyri were primarily driven by pictorial stimuli and mediated by bottom up processing. Nevertheless, there are reports of similar category effects when the stimuli are written names (Chao *et al.,* 1999; Devlin *et al.*, 2005). This suggests that the effects are not entirely driven by the perceptual input. One interpretation of the medial fusiform category effects is that activation reflects the semantic relevance of an object's visual features (see Mechelli *et al.*, 2006). Semantic relevance is a measure of the distinctiveness and importance of an object's features, that is, concepts may have many semantic features but only a small number of features that are relevant for distinguishing it from closely related concepts. For example, "trunk" is a feature with high semantic relevance for the concept "elephant", whereas "has four legs" is of low semantic relevance for defining the concept "elephant" because this feature is used to define many (both living and non-living) concepts.

To recap, right medial fusiform activation at [30, -42, -24] was higher for nonverbal than verbal audiovisual matching even when visual input was controlled. The same regions are sensitive to object category and semantic relevance. In Mechelli *et al.* (2006), peak activation for pictures of artefacts relative to animals was located at [28, -42, -22] and [-28, -52, -14] which is less than 3mm away from the nonverbal audiovisual integration effects in the right fusiform. Likewise, Mechelli *et al.* (2006) located the effect of semantic relevance in close proximity at [30, -46, -14]. Might there be a single underlying function that accounts for all these effects? The strong effects of object category and semantic relevance during picture processing could simply reflect bottom-up structural object processing (Noppeney *et al.,* 2006). As shown in Figure 6.4 and Table 6.5, the nonverbal effect during audiovisual processing is primarily driven by picture stimuli. Nevertheless, purely nonverbal stimuli (pictures with sounds) increased right medial activation, all be it weakly, relative to picture stimuli that were presented with spoken words (see Table 6.5). One possibility is that the demands on bottom up structural processing are reduced in the presence of spoken words than environmental sounds. This is consistent with response times being more than 100ms faster for the AwVp versus AsVp stimuli (see Table 6.1). However, faster response times and reduced structural processing could simply reflect the fact that spoken words were recognised before environmental sounds thereby speeding up the matching process. An alternative explanation of the nonverbal effect is that the increased right medial fusiform activation for pictures with sounds relative to pictures with spoken words might reflect access to stored knowledge of object structure. This would be consistent with the "semantic processing" conclusions of Chao *et al.* (1999) following their observation of category effects in these regions during written word processing. Clearly further experiments are required to test the structural and semantic memory hypotheses. For example, the structural processing hypothesis could be tested with an experiment that manipulated the timing of the auditory stimuli so that environmental sounds were recognised at the same time point as spoken words. If medial fusiform activation was still higher for the purely nonverbal stimuli, this would be more consistent with the semantic than perceptual hypothesis. Conversely, if the nonverbal effect was lost, this would be more consistent with the structural processing hypothesis.

6.6 Summary

In conclusion, the results of this chapter comparing verbal and nonverbal audiovisual matching confirm the prediction that purely verbal audiovisual matching increases phonological activation relative to conditions with one or no verbal component. A second expectation was that nonverbal stimuli would be more reliant on semantic processing areas. However, contrary to this prediction, the right medial fusiform gyrus was the only region to show increased activation for nonverbal relative to verbal stimuli. Although the medial fusiform gyri have been associated with stored structural knowledge and semantic relevance, further experiments are required to exclude confounds from the impact of temporal differences in spoken word and sound identification. The most robust conclusion of this chapter is therefore the observation that purely verbal audiovisual stimuli can be matched faster than nonverbal stimuli with increased activation in a left STS region associated with phonological processing. Critically, this STS area [-62, -36, 4] does not correspond to that investigated in the previous chapters [+/-50, -52, 8] and [+/-50, -56, 4].

Chapter 7

7 Crossmodal integration of tactile-visual abstract shapes

7.1 Introduction

The neural substrate for crossmodal matching between the tactile and visual modalities in humans is not well understood. To date there have been only a small number of functional imaging studies on this topic published, with a wide variation in the experimental paradigms utilised. The lack of consistency across stimulus and task parameters in these studies has correspondingly produced a lack of consistency in candidate regions for the integration of tactile and visual signals (see Amedi *et al.*, 2005 for a review). In contrast, data from non-human primate studies have provided more consistent data. The primary focus has been the effect that medial temporal lobe lesions have on the processing of tactile-visual inputs (Goulet and Murray, 2001; Murray and Bussey, 1999; Murray and Richmond, 2001), although the underlying functional processes that drive normal activation, and that are impaired following lesions of this region, are still a matter of debate (see Buckley and Gaffan, 2006; Bussey and Saksida, 2005 for reviews). Functional deficits following brain damage in human patients has also found some support for a role of the medial temporal regions in the processing of combined tactile-visual inputs (Shaw *et al.*, 1990). The experiment reported here investigates the neural regions engaged during tactile-visual object processing in normal subjects, in the context of controlling the attentional, task and stimulus confounds highlighted in the previous chapters.

7.11 Findings from non-human primate studies

The effect of rhinal cortex ablation on non-human primates has demonstrated a critical role for the entorhinal and perirhinal cortices in the formation of crossmodal tactile-visual representations. Initial studies demonstrated that monkeys lost the ability to learn crossmodal tactile-visual associations with food (Cowey and Weiskrantz, 1975; Jarvis and Ettlinger, 1977) and meaningless objects (Goulet and Murray, 2001; Murray and Mishkin, 1985). For instance, on tests of delayed non-matching-to-sample (DNMS), consistent impairments were observed during crossmodal performance (Goulet and Murray, 2001). However, performance also appears to depend on the perceptual difficulty of the task (Buckley and Gaffan, 1997; Eacott *et al.,* 1994). This has recently led to the view that perirhinal cortex is involved not just in mnemonic processes related to learning, but also has a role in the perceptual processing of objects (Murray and Gaffan, 1994; Murray and Bussey, 1999; Murray and Richmond, 2001; Parker and Gaffan, 1998). Interestingly, this region of the ventro-medial temporal lobe receives a wide variety of sensory inputs, including somatosensory afferents via the insula (Friedman *et al.,* 1986; Suzuki, 1996; Suzuki and Amaral, 1994a), and is situated at the anterior end of the ventral visual processing stream (Murray and Bussey, 1999), making it well placed to processes simultaneous multi-sensory inputs.

7.12 Neuropsychological data in humans

Data from human patients with brain-damage have also implicated the medial temporal lobes in crossmodal tactile-visual processing. For example, Shaw *et al.* (1990) tested patients suffering from herpes simplex virus encephalitis (HSVE) on a crossmodal versus intra-modal tactile-visual perceptual discrimination task using simultaneously presented simple arcs and circles. The patients were able to match within the same modality, but were impaired on the crossmodal tactile-visual matching task. Similarly, data from two further patients (unpublished data summarised in Holdstock, 2005) have also implicated medial temporal lobe damage in the observed impairment for crossmodal tactile-visual matching relative to intra-modal visual or tactile tasks. Despite the presence of a dissociation between intra- and crossmodal matching, lack of information on the specific location and extent of brain damage make studies such as these difficult to interpret in terms of the neural regions normally engaged by crossmodal tactile-visual integration.

7.13 Functional imaging of normal subjects

In contrast to the apparent consistency between lesion data from non-human primates and deficits observed in patients with medial temporal lobe damage, activation in this medial temporal region has not been specifically investigated, or reported, in functional imaging studies of tactile-visual matching. On the contrary, a review of the literature by Amedi *et al.* (2005) found three alternative regions implicated in the functional integration of tactile-visual forms: 1) the lateral-occipital complex (in particular LOtv, Amedi *et al.*, 2001), 2) the intra-parietal sulcus, both reportedly involved in object shape analysis and recognition, and 3) the insula/claustrum, suggested to be more involved in the crossmodal binding of this sensory information. Only one of the studies reviewed by Amedi *et al.* (2005) reported activation in the medial temporal lobe (specifically the para-hippocampal gyrus; Stoesz *et al.*, 2003), however this experiment used tactile discrimination only, with no visual stimulus input. It is therefore not possible to conclude from that particular study that the region is involved in the integration of tactile-visual inputs.

7.14 This Study

From this brief summary of the literature, the discrepancy between the findings from non-human primates and results of functional imaging of normal subjects appears intriguing. In chapters 4 and 5, I suggested a number of methodological factors related to the design of functional imaging studies that may contribute to these differences. These include 1) the inconsistent level of sensory input between uni-modal and crossmodal stimuli, 2) the effect that a passive versus active task has on regional brain activation, and 3) the differences between processing congruent (referring to one object) versus incongruent (referring to two objects) sensory inputs. In this chapter fMRI was used to investigate the neural substrates of intra-modal versus crossmodal tactile-visual shape matching. Three experimental conditions manipulated the sensory modality of abstract object pairs, comprising either intra-modal visual-visual matching, intra-modal tactile-tactile matching or crossmodal tactile-visual matching, with a match/no-match decision task. To control for the aforementioned potential confounds, tactile and visual stimulus input was kept constant across all intra-modal and crossmodal conditions, and an active matching task was also held constant. In addition the data were analysed in an event-related fashion, in the context of a blocked design, in order to control for attentional differences that would be unavoidable during blocks that used *only* congruent (matching object concepts) or incongruent (different object concepts) trials. In this way, the study was able to investigate the brain regions involved in crossmodal relative to intra-modal matching of tactile-visual abstract shapes, and determine whether this activation was modulated by congruency (i.e. matching versus non-matching object concepts).

7.2 Methods

7.21 Materials and Methods

Subjects

18 subjects (5 female, 13 male, mean age 21.3) took part in the study. All were right handed native English speakers with normal or corrected to normal vision and gave informed consent to take part. All had normal neurological status. The study was approved by the joint ethics committee of the Institute of Neurology and University College London Hospital, London, UK.

Experimental design

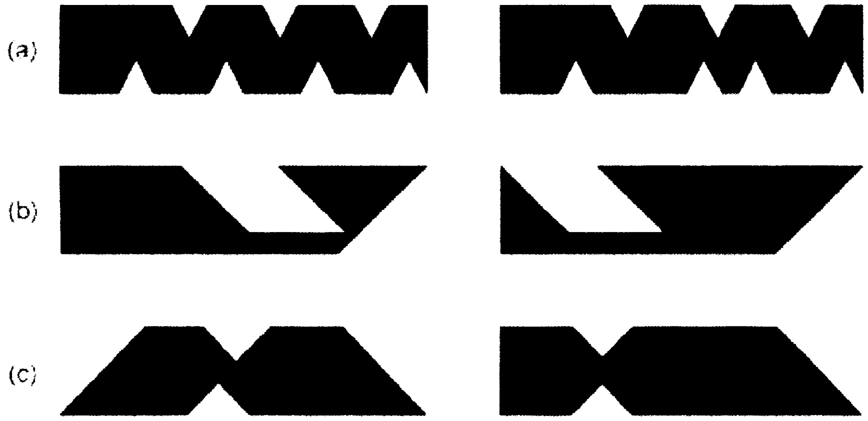
In this experiment, subjects were instructed to indicate, with a left or right foot movement, whether two simultaneously presented meaningless shapes were identical or not. There were two types of stimuli: visual (a silhouette abstract shape or a silhouette circle), and tactile (2-dimensional wooden blocks or wooden spheres). In visual-visual (VV) matching conditions, two shapes were presented, on a back-projected screen, either side of a fixation cross. During tactile-tactile (TT) matching, two wooden shapes were explored with both hands simultaneously. In crossmodal trials (TV), the tactile stimuli were presented to either the left or the right hand and the visual stimuli were presented either left or right of a central fixation point, balancing visual presentation (left or right side of screen) and palpating hand (left or right). An additional baseline condition presented silhouette circles (visual) and wooden spheres (tactile), where subjects made alternating foot movements for each trial (i.e. no match decision was required). The order of these conditions was blocked and counterbalanced between and within subjects. Within each block, there were three congruent trials (both stimuli referred to the same shape, requiring a Yes-they match response) and three incongruent trials (stimuli referred to two different shapes, requiring a No-they do not match response). The order of the congruent and incongruent trials was randomised within block.

To control for perceptual (visual and tactile) input across conditions, subjects were also presented with silhouette circles or wooden spheres on the visual screen or hand that was not engaged by the matching task. For example, during VV matching, subjects manipulated a wooden sphere in both hands with presentation rate corresponding to the rate of visual shape presentation. For TT matching, subjects viewed two circles presented either side of the fixation point at the same rate as the wooden abstract shapes were presented to their hands. Finally, during TV matching, meaningless spheres/circles were presented to both the side of screen and hand that were not engaged with the crossmodal shape matching task. Thus in all conditions subjects received simultaneous tactile and visual stimulation, manipulating either wooden shapes or spheres in both hands and viewing either visual shapes or silhouette circles, depending on the experimental condition.

Stimuli

Each visual or tactile stimulus was an abstract shape created by removing small rectangles and triangles from an original rectangle or a circle/sphere (see Figure 7.1). Visual stimuli were abstract silhouette shapes (black) or a silhouette circle. Tactile stimuli were made from a wooden block (overall dimensions: 10cm x 2.5cm x 1.2cm) formed into a 2-dimensional abstract shape. Wooden spheres were used during tactile control conditions. These tactile stimuli were mounted onto card to facilitate presentation by the experimenters during scanning, and to ensure that wooden blocks were always presented at the same angle as the visual stimuli. In total there were 72 visual and 72 tactile shapes. To equate task difficulty across matching conditions, and thereby ensure that neural activation was not confounded by differential task demands between the three conditions, visual matching shapes in the VV condition contained more features than those in the TT and TV conditions.

Figure 7.1: Sample stimuli used during each matching condition



(a) A non-identical stimulus pair from the VV matching task

(b) Two dimensional shapes of a non-identical stimulus pair from the TT matching task

(c) Two dimensional shapes of a non-identical stimulus pair from the TV matching task

Procedure

All subjects were familiarised with the stimuli and trained on the task prior to scanning. In addition, one training session containing all condition types was carried out with subjects lying in the scanner, without acquiring data, in order for subjects to be fully practiced on the task, in particular to co-ordinate presentation of the tactile stimuli with the experimenters. For each block, stimuli were presented for 3.6s followed by 0.63s fixation to allow for the response, resulting in a total block time of 25.38s. Each block of 6 trials was followed by 16.2s of fixation and a 3.78s period during which visual instructions were displayed. Over the experiment, each subject was presented with 32 crossmodal blocks (96 congruent and 96 incongruent trials) plus 8 blocks of intra-modal visual and 8 blocks of intra-modal tactile matching. There were four different sessions with eight blocks of crossmodal tactile-visual matching [two for each combination of: 1) right hand-left side of screen; 2) left hand-right side of screen; 3) right hand-right side of screen; 4) left hand-left side of screen] counterbalanced using an ABBA design. The presentation of the two tactile stimuli was controlled by two experimenters on either side of the subject. Each experimenter received auditory cues via headphones to indicate the presentation time that coincided with the onset of the visual stimuli. Subjects were instructed to palpate the circles/spheres and decide whether the tactile and visual silhouettes had the same shape or not.

7.22 Data Acquisition

Data were acquired on a Siemens 1.5 Tesla scanner (Siemens, Erlangen, Germany). Functional images used a T2*-weighted echo-planar (EPI) sequence for BOLD contrast with 3 x 3 mm in plane resolution, 2mm slice thickness and a 1mm slice interval. 36 slices were collected resulting in an effective repetition time (TR) of 3.24sec/volume). After the functional sessions, a T1-weighted anatomical volume image was acquired from all subjects to ensure normal neurological status.

Data Analysis

Functional data were analysed with statistical parametric mapping (SPM2, Wellcome Trust Centre for Neuroimaging, London, UK) implemented in Matlab. 7.1 (Mathworks, Sherborne, MA, USA). Pre-processing included realignment and unwarping using the first volume as the reference scan (after excluding the first 4 dummy scans to allow for T1 equilibration effects) spatial normalisation to a standard MNI template (Friston *et al.,* 1995a) and spatial smoothing using a 6mm full width half maximum isotropic Gaussian kernel. One subject was removed from the analysis due to excess head movement.

First level statistical analyses (single subject and fixed effects) modelled each trial type independently by convolving the onset times with the haemodynamic response function. There were 12 different trial types: 4 intra-modal conditions [visual or tactile matching] x [congruent or incongruent], and 8 crossmodal conditions [tactile-visual matching with left or right visual presentation on screen and left or right palpating hand] x [2 congruent or incongruent]. The data were high-pass filtered using a set of discrete cosine basis functions with a cut-off period of 128sec. Parameter estimates were calculated for all voxels using the general linear model, by computing a contrast image for each trial type relative to the baseline condition. The parameter estimates were then fed into a second level ANOVA. Unless stated otherwise, all significant effects are reported at $p<0.05$ corrected for multiple comparisons across the whole brain.

7.3 Behavioural Results

A 2x2 repeated measures ANOVA with factors of crossmodal vs. intra-modal matching and congruent versus incongruent trials was conducted on the subject responses. There were significantly more correct responses for crossmodal than intra-modal matching (F(1,16)=9.907, $p<0.05$) and for congruent than incongruent trials (F(1,16)=7.761, $p<0.05$) but no interaction between modality and congruency (F(1,16)=2.589, $p>0.05$). Means and standard deviation for each condition are show in Table 7.1.

Table 7.1: Response accuracy for tactile-visual matching

Matching Condition	*Percent correct (SD) congruent trials*	*Percent correct (SD) incongruent trials*	*Percent correct (SD) all trials*
Intra-modal matching			
Tactile-tactile	79.7 (11.1)	83.8 (16.7)	82.0 (15.0)
Visual-visual	96.1 (5.4)	84.1 (13.8)	89.7 (11.3)
Mean	*87.9 (5.9)*	*84 (11.5)*	*85.9 (10.0)*
Crossmodal matching			
Tactile-visual left/left	96.1 (7.6)	92.7 (6.8)	94.4 (8.5)
Tactile-visual left-right	96.1 (6.1)	82.6 (12.7)	89.0 (10.7)
Tactile-visual right/right	94.8 (5.4)	89.5 (12.2)	92.2 (9.4)
Tactile-visual right/left	92.8 (7.0)	83.0 (12.2)	87.9 (12.3)
Mean	*94.7 (4.7)*	*86.9 (9.8)*	*90.8 (7.8)*

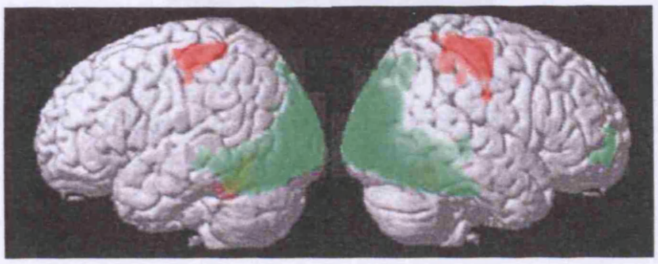
Means and standard deviation (SD) in parentheses for correct responses during: 1) congruent matching 2) incongruent matching and 3) all matching trials.

7.4 Functional Imaging analysis

7.41 Intra-modal matching

Matching tactile pairs activated bilateral superior parietal cortices relative to the TV and VV conditions, and visual pairs activated bilateral occipito-temporal cortices relative to TV and TT conditions (see Figure 7.2).

Figure 7.2: Main effect of intra-modal matching



Rendered on the SPM averaged MNI template, with TT > TV +VV in red and VV > TV + TT in green, at a threshold of p<0.001 uncorrected.

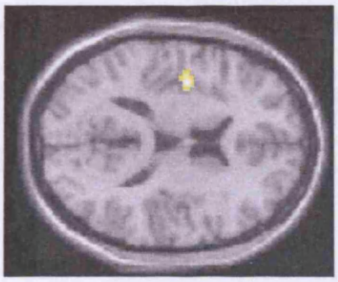
7.42 Crossmodal versus intra-modal matching

Across the whole brain, one cluster in the left insula reached a corrected level of significance by extent for all crossmodal pairs relative to all intra-modal conditions [-38, -6, 14; Z=4.7, k=135, see Figure 7.3].

7.43 Congruent versus incongruent matching

Across all conditions, no regions were more activated by congruent relative to incongruent trials, but matching incongruent relative to congruent trials increased activation in the right posterior middle frontal gyrus [48, 14, 36]. This effect was significant in extent (k=101) but not height (Z=4.6), after correction for multiple comparisons across the whole brain. There were no significant interactions between congruency and crossmodal relative to intra-modal matching.

Figure 7.3: Crossmodal versus intra-modal matching



Insula activation with the cluster peak at [-38, -6, 14] rendered at $p<0.001$ uncorrected, shown in MNI space on an averaged T1 structural image in the axial plane at $z = 14$.

7.5 Discussion

Matching crossmodal relative to intra-modal pairs increased activation in the insula, independent of congruency. These data are particularly interesting in that they replicate previous findings implicating the insula in crossmodal integration (Calvert, 2001), although there is no consistency for lateralisation of these effects in tactile-visual matching: Banati *et al.* (2000) report left insula, whereas Hadjikhani and Roland (1998) report activation in the right insula.

A review of the crossmodal literature by Amedi *et al.* (2005) implicated two additional regions in the functional integration of tactile-visual forms: 1) the lateral-occipital complex (in particular LOtv), and 2) the intra-parietal sulcus (IPS). In addition, lesion data from neuropsychological patients and non-human primates has proposed a role for the medial temporal lobes in tactile-visual integration. No activation was significant in these regions after a correction for multiple comparisons across the whole brain, and no regions were differentially involved in congruent versus incongruent matching at a height-level corrected threshold. The following discussion will firstly examine the functional role that the insula is playing in the integration of these crossmodal signals. I will then turn to the function of the three alternative cortical regions implicated in previous crossmodal studies to determine why and how the data here differ from previous findings.

7.51 Crossmodal versus intra-modal matching in the insula

Consistent with two previous studies (Banati *et al.,* 2000; Hadjikhani and Roland, 1998) the insula was more activated during tactile-visual matching that required crossmodal integration of information about the shapes of objects, than intra-modal matching that could be achieved by processing within uni-modal regions. This finding provides support for the proposal that the insula acts as a mediating region that enables communication and exchange of information between uni-modal regions (Amedi *et al.,* 2005; Calvert, 2001). Moreover, the current study found equivalent activation in the insula for congruent and incongruent trials, i.e. irrespective of whether or not the visual and tactile information could be successfully integrated into a unitary object representation.

Activation in the insula has been found in other studies involving somatosensory processing. Downar *et al.* (2000) reported activation in right posterior and anterior insula (in addition to other regions) during passive detection of transitions between simultaneous multi-modal tactile, visual and auditory stimulation and periods of no stimulation. In a task using tactile stimuli only, Reed *et al.* (2004) reported activation in the left insula during object recognition relative to tactile non-objects. Although there was no visual stimulus in their experiment, subjects were required to name the tactile object silently whereas the non-object palpation did not have a task. It is therefore possible that subjects were engaging in visual imagery of the real objects, with the insula (in addition to other reported regions) mediating between the uni-modal tactile and visual association cortices.

7.52 Integration in intraparietal sulcus

Other studies of tactile-visual integration have reported activation in the IPS for tactile-visual processing (Banati *et al.,* 2000; Grefkes *et al.,* 2002; Saito *et al.,* 2003). For instance, Saito *et al.* (2003) reported posterior IPS but no insula activation in their crossmodal relative to intra-modal matching task, although the task was quite different from that used here (exploring complex patterns with the thumb only relative to the 3-dimensional objects with all fingers used here). In their task, Saito and colleagues controlled stimulus intensity across the crossmodal and intra-modal matching conditions. However, they did not control for difficulty across task, with behavioural data showing that the intra-modal visual matching condition was significantly easier than the other conditions. The increased task demands placed on matching tactile-tactile and tactile-visual input may therefore have differentially weighted activation in this region toward tactile processing.

The inferior parietal lobe activations reported to be associated with tactile-visual processing by Banati *et al.* (2000) may also reflect tactile processing rather than tactile-visual integration (Banati *et al.,* 2000; Deibert *et al.,* 1999). Banati and colleagues compared activation during tactile-visual matching with that during visual-visual matching only. As a result, the activations identified in their subtraction analysis reflect both tactile processing and tactile-visual integration. Accordingly, the IPS activation reported in Saito *et al.* (2003) during their crossmodal relative to intra-modal matching task could have been due to an enhanced response to processing tactile inputs.

Grefkes *et al.* (2002) reported left anterior IPS activation during a crossmodal delayed matching to sample task. Subjects were required to transfer visually learned information for a tactile response and learn tactile information for a visual response, so information was transferred between modalities for matching, rather than integrated into a perceptual whole. The design also meant that there was no simultaneous presentation of visual and tactile stimuli, as used here. The anterior IPS region they reported has previously been associated with object grasping (Binkofski *et al.,* 1998), manipulation of complex objects (Binkofski *et al.,* 1999) and manual shape discrimination (Bodegard *et al.*, 2001). The question then is whether tactile processing places greater demands on this region, and that activation reported in these studies reflects these increased demands. This question was explored in the set of data reported here. Using a lower threshold ($p<0.001$ uncorrected) for crossmodal versus intra-modal matching, with their co-ordinates [-40, -42, 36] as a region of interest (with a small volume correction sphered at 10mm) revealed a peak voxel located at [-42, -34, 38; Z=4.2]. This effect was driven by tactile-visual and tactile-tactile matching only, with no effect for visual-visual matching. (NB this is also close to the peak of a large cluster in the superior parietal lobe reported by Hadjikhani and Roland, 1998 for the intersection of TT-control and TV-VV with peak co-ordinates at [-44, -31, 47]). In summary, although IPS activation has been observed in a small number of studies, it seems likely that this activation is driven by the demands placed on processing tactile input, rather than being a region critical for the crossmodal processing of tactile-visual inputs.

7.53 Integration in LOtv

In a review of multi-sensory integration by Amedi *et al.* (2005) the lateral-occipital complex (in particular LOtv, Amedi *et al.,* 2001) and the IPS were found to be involved in tactile-visual processing, in addition to the insula. These regions were reportedly involved in object shape analysis and recognition, whereas a review of the function of the insula/claustrum suggested it was more involved in the crossmodal binding of this sensory information. In the present study, activation in the lateral occipital cortex or IPS was not found for crossmodal relative to intra-modal matching (at a level corrected for multiple comparisons across the whole brain). Moreover, previous studies that have specifically engaged subjects in a crossmodal relative to intra-modal matching task have also not found activation in these regions.

Of those studies reviewed by Amedi *et al.* (2005), the prime focus was on studies that found engagement of visual regions during tactile processing, rather than integration or matching across modalities *per se*. Tasks where LOtv was involved included: uni-modal visual object and somatosensory object recognition (Amedi *et al.,* 2001, 2002); crossmodal tactile-visual priming (James *et al.,* 2002); micro- versus macro-tactile form discrimination (Stoesz *et al.,* 2003); tactile real versus tactile abstract objects (Reed *et al.,* 2004); visual and tactile (but not concurrent) recognition of faces and real objects (Pietrini *et al.,* 2004); and discrimination of tactile form versus orientation using mental rotation (Prather *et al.,* 2004). Although these studies clearly demonstrate the involvement of LOtv in both visual and tactile processing of real and non-objects, their design does not permit the conclusion that these regions are involved in crossmodal relative to intra-modal matching of simultaneous tactile-visual inputs. Rather they demonstrate that LOtv is responsive to both tactile and visual inputs during a range of tasks. The findings reported here indicate that LOtv is not involved in the crossmodal binding of tactile-visual inputs, however it cannot speak to the issue of the engagement of LOtv in the absence of visual input, as both tactile and visual sensory input were consistent across all conditions.

7.54 Integration in medial temporal lobes

In the non-human primate literature, lesions to rhinal cortices have been found to disrupt crossmodal tactile-visual associations (e.g. Goulet and Murray, 2001). This is in contrast to the insula activation which is proposed to be facilitating communication between uni-modal regions (Amedi *et al.*, 2005). The functional role of the medial temporal lobes is debated in the literature, with a question over its involvement in purely mnemonic processes, or whether it is also engaged by perceptual processing (Murray and Bussey, 1999; Murray and Richmond, 2001). The crossmodal stimuli used here are meaningless objects, which would have no stored object representation (Murray and Bussey, 1999; Murray and Richmond, 2001). These data would then appear to support the view that the medial temporal regions are not engaged by the perceptual abstract stimuli used here. However, an ever-increasing literature places a central role for the medial temporal regions, in particular perirhinal cortex, in the integration of crossmodal inputs (e.g. Taylor *et al.*, 2006). Moreover, the consistency of findings from alternative sources suggests that medial temporal cortex *is* involved in crossmodal integration. The following chapter has therefore been designed to look at these regions in more detail, using a region of interest approach, to look specifically at the role of perceptual versus mnemonic processing.

7.6 Summary

These results critically replicate the finding of activation in the insula for crossmodal tactile-visual relative to intra-modal matching of tactile or visual abstract shapes. This is consistent with previous studies that have specifically investigated cross-relative to uni-modal integration (Banati *et al.,* 2000; Hadjikhani and Roland, 1998). However, contrary to previous reports, there was no increased activation in the IPS or LOtv at a level of significance corrected for the whole brain. Although it is difficult to compare results across studies that use different tasks, the data suggest that activation observed in IPS is the result of increased demands on processing in the tactile domain, independent of whether these are within or between input modality. Based on previous findings, activation in LOtv appears to be engaged when processing 1) meaningful multi-modal objects, rather than crossmodal versus intra-modal processing *per se*, and 2) tactile processing when visual imagery or access to visual representations of real objects is required.

Based on findings from non-human primate and neuropsychological data it is perhaps surprising that activation was not observed in the medial temporal lobes. Would activation in this region be driven only by amodal representations? There is no data in this thesis investigating meaningful tactile-visual object processing, hence no possibility of directly comparing meaningless tactile-visual shapes versus meaningful tactile-visual concepts. However, the same design was used for crossmodal versus intra-modal audiovisual object processing reported in Chapter 4. The following chapter combines the data from the crossmodal conditions of these two experiments to further investigate multi-modal congruency and meaning in object processing, with particular focus on the role of anterior and medial temporal lobe regions in processing perceptual versus semantic inputs.

Chapter 8

8 Crossmodal object processing in the lateral and medial anterior temporal cortex

8.1 Introduction

The critical contribution of the anterior temporal lobes to object processing and representation is clearly reflected in the behavioural deficits of patients with damage to these regions due to disease such as semantic dementia (SD). Anatomically, SD is characterised by significant atrophy of the anterior temporal poles (usually bilaterally), progressing caudally into infero-lateral and medial regions of the temporal cortex. In behavioural terms, this group of patients have a generalised semantic deficit, performing poorly on tasks requiring access to conceptual knowledge, independent of the modality of presentation, i.e. spoken/written linguistic information, picture processing or environmental sound input (e.g. Bozeat *et al.,* 2000; Garrard and Hodges, 2000; Lambon-Ralph *et al.,* 2001; Rogers *et al.,* 2004, 2006). As the disease progresses, their behavioural performance deteriorates: for example, in object naming tasks, responses become progressively less specific, such as using the name dog instead of Labrador, or animal rather than dog.

Patients with herpes simplex virus encephalitis (HSVE) also present with semantic problems, but without the all-encompassing deficits that are observed in SD. Interestingly, atrophy is predominantly to medial temporal lobes, in the context of relative sparing of the antero-lateral temporal cortex (Gainotti, 2000; Noppeney *et al.,* 2007b). In a comparison between the semantic processing differences of SD and HSVE patients, Levy *et al.* (2004), reported that HSVE patients scored better than

SD on standardised neuropsychological tests of semantic knowledge. Interestingly, the patient with HSVE who was most impaired on tasks of semantic knowledge had the most damage to antero-lateral temporal cortex.

Until now, focus has been placed on the role of the medial temporal lobes in object processing and representation, with some debate on whether the role of this region is in perceptual or semantic processing (Buckley & Gaffan, 2006; Bussey & Saksida, 2005; Murray *et al.,* 2005). Convergent evidence from both the human and non-human primate literature increasingly suggests however that this antero-medial temporal region, in particular the perirhinal cortex, is involved in both access to semantic representations and differentiating between perceptually similar items in tasks with no memorial component (Bussey *et al.* 2003; Lee *et al.*, 2005, 2006).

An unanswered question is what function the antero-lateral regions of the temporal lobes play in the representation and processing of semantic knowledge. A recent functional imaging study by Rogers *et al.* (2006) reported antero-lateral temporal activation in a task where normal subjects identified concepts at a specific relative to general level. Interestingly, this region corresponded to the site of maximal atrophy in patients with SD (Mummery *et al.,* 2000). In the context of supporting data, Rogers *et al.* (2006) reasoned that this region was involved in crossmodal semantic, rather than purely visual perceptual, processing.

Taken together, these findings suggest the possibility that different functional processes are engaged by anterior infero-lateral and medial temporal regions. The more selective deficit for knowledge of living things often observed in HSVE patients (see Capitani *et al.,* 2003 for a review) suggests that the medial temporal lobes are critically involved in perceptual discrimination, with relatively successful access to stored representations when perceptual object discrimination is less demanding, as reflected in their relatively preserved ability to name non-living items. In contrast, the generalised semantic deficit in SD patients and the lack of category specificity suggests that their deficit could be the result of more widespread damage, including anterior infero-lateral cortex, which impairs their access to stored representations as well as perceptual discrimination impaired as a result of antero-medial temporal atrophy.

This study

Here I investigate whether placing different demands on object discrimination processes in normal subjects engages one or more of these anterior temporal regions differentially. As proposed in the preceding chapter, the complementary nature of the experimental design for tactile-visual (Chapter 7) and audiovisual (Chapter 4) matching enabled a second-level comparison between activation in both types of crossmodal trials. This allowed further investigation of multi-modal congruency and meaning in object processing, to address questions raised about the functional role of anterior temporal regions in normal object processing, based on findings from the human neuropsychological and non-human primate literature. To summarise, employing a crossmodal decision task, subjects had to decide whether two crossmodal stimuli were either congruent (i.e. they refer to the same object) or incongruent (i.e. they represent two different objects). Increased activation for congruent trials suggests a role in crossmodal integration, whereas increased activation for incongruent trials would be expected with increased perceptual/conceptual information processing (i.e. two perceptual/conceptual inputs per trial). In addition, there were two different kinds of crossmodal matching: one experiment used an audiovisual conceptual matching task and the other experiment used tactile-visual perceptual matching. Audiovisual trials consisted of meaningful objects, whereas tactile-visual trials used meaningless abstract shapes. Comparing the effects of these experiments therefore investigated different levels of object processing in the anterior temporal regions that are damaged in SD.

8.2 Method

8.21 Materials and Methods

Subjects

There were 18 subjects (11 women, 7 men, mean age 26) in the audiovisual experiment (Experiment 1) and 18 subjects (5 female, 13 male, mean age 21.3) in the tactile-visual experiment (Experiment 2). All were right handed native English speakers with normal or corrected to normal vision, and gave informed consent to take part. The study was approved by the joint ethics committee of the Institute of Neurology and University College London Hospital, London, UK.

Experimental design and stimuli

Audiovisual matching

Full details on the stimulus, design and acquisition parameters for the audiovisual conceptual matching experiment can be found in Chapter 4. To briefly recap, the experiment used four types of object stimuli: pictures of objects, their written names, their auditory names and their associated environmental sounds (i.e. two visual and two auditory). This resulted in four different conditions, with an equal number of congruent and incongruent trials within each condition:

(1) Crossmodal trials with pictures of objects and spoken names,

(2) Cross modal trials with written names and environmental sounds.

(3) Intra-modal visual trials with one picture and one written name, and

(4) Intra-modal auditory trials with one spoken object name and one environmental sound.

Subjects made a match or no match decision on each trial using a key press response. For the experiment reported here, the contrast of interest was the difference between activation for crossmodal congruent and incongruent conditions (conditions 1 and 2 combined).

Tactile-visual matching

Full details on the tactile-visual conceptual matching experiment can be found in Chapter 7. As a reminder, in that experiment stimuli were visual (a silhouette abstract shape or a silhouette circle), and tactile (2-dimensional wooden blocks or wooden spheres). This gave a total of four crossmodal (tactile-visual matching with left or right visual presentation on screen and left or right palpating hand) and 2 intra-modal conditions (visual or tactile matching). These conditions contained an equal number of congruent and incongruent trials. For the comparison reported in this chapter, the contrast of interest was the difference between congruent and incongruent conditions involving crossmodal trials only, independent of the side of presentation (left or right). Note that although the stimulus duration and presentation rate was different for the audiovisual and tactile-visual experiments (see Chapters 4 and 7 for details), these parameters were matched within experiment for congruent and incongruent stimuli.

8.22 Data Analysis

At the first level, parameter estimates were calculated for all voxels using the general linear model, by computing one contrast image for congruent relative to incongruent crossmodal trials for each experiment. These parameter estimates were then fed into a second level ANOVA that modelled the effect of congruent vs. incongruent in (a) the combined audiovisual (AV) crossmodal conditions, and (b) the combined tactile-visual (TV) crossmodal conditions. This allowed testing of the main effect of congruency (congruent vs. incongruent) and its interaction with experiment (AV vs. TV). Note that differences between experimental procedures in the two experiments are controlled by limiting inter-experiment comparisons to the interaction with the effect of congruency. The *t*-images for each contrast at the second level were subsequently transformed into the statistical parametric maps of the Z statistic. Unless stated otherwise, all significant effects are reported at $p<0.05$ corrected for multiple comparisons across the whole brain or in regions of interest (detailed below).

Regions of Interest

The evidence reviewed in the Introduction of this chapter suggested that different regions of the anterior temporal lobes are engaged by different functional processes. These differences are based on behavioural deficits observed in patients with SD versus those with HSVE. To further investigate different levels of object processing in the anterior temporal regions that are damaged in semantic dementia, a region of interest (ROI) was used. This ROI was centred (with 12mm radius) on the co-ordinates of peak atrophy [+/-44, 14, -27] in patients with SD reported in Mummery *et al.* (2000; see Rogers *et al.*, 2006 for details of the region of interest). A second ROI was located in the antero-medial temporal lobe, where debate continues as to the functional role of this region in perceptual versus mnemonic object processing. Using the co-ordinates reported by Taylor *et al.* (2006) for the effect of incongruent crossmodal conceptual matching [-26 -20 -22], this ROI was used to determine whether there are differences within this region between perceptual (i.e. non-meaningful) and conceptual crossmodal matching.

8.3 Behavioural results

In all AV trials, subjects made only a small number of errors (congruent trials: 1.1%, incongruent trials 2.4%). In comparison, across all TV trials, incorrect responses for congruent trials were 4.7% and 9.8% for incongruent.

8.4 Functional Imaging Results

8.41 Congruent > Incongruent Trials (over experiments)

Matching congruent versus incongruent crossmodal pairs increased activation in the medial fronto-polar gyrus and the precuneus. In the regions of interest there was increased activation in the left anterior temporal pole, the peak co-ordinates of which were only a few millimetres away from the centre of the area damaged in SD. See Table 8.1 and Figure 8.1a for details.

Table 8.1: Main effect of congruency

Contrast	Anatomical Region	Co-ordinates			Z-score
		x	y	z	
Congruent > Incongruent	Right medial fronto-polar gyrus	4	60	4	5.1
	Left Precuneus	-2	-70	38	5.0
	Left anterior temporal pole	**-44**	**12**	**-36**	**4.3**
Incongruent > Congruent	Left posterior occipito-temporal	-44	-70	-14	4.4

Table gives anatomical regions, co-ordinates and Z-scores for the effect of congruency across Experiments 1 and 2. Effect in the region of interest is shown in bold.

8.42 Incongruent > Congruent Trials (over experiments)

Matching incongruent versus congruent crossmodal pairs increased activation in the left inferior occipital gyrus (see Table 8.1). This effect was significant in extent (k=115) but not height (Z=4.4). There were no significant effects in the regions of interest.

8.43 Interaction between congruency and experiment

For the AV experiment only, incongruent relative to congruent trials increased activation in two regions of the antero-lateral inferior temporal lobe. Both effects were qualified by a congruency by experiment interaction (see Table 8.2) which survived correction for multiple comparisons across the whole brain. Of interest here is that the location of the incongruency effects during audiovisual object matching were within the area of "typical" atrophy in SD (see Figure 8.1b) but posterior to the amodal congruency effect reported above.

There were no significant effects of congruency that were greater in the TV shape matching experiment than the AV object matching experiment in the region of interest. However, in the medial temporal area that Taylor *et al.* (2006) reported for incongruent audiovisual relative to the uni-modal stimuli [-26, -20, -22], there was a trend for higher activation for incongruent > congruent in the AV experiment and the reverse trend (congruent > incongruent) in the TV experiment. The interaction between congruency and experiment was significant using a small volume correction centred on the co-ordinates [-26, -20, -22] from the Taylor study with a 12mm radius. See Table 8.2 and Figure 8.1c for details.

Table 8.2: Interaction between congruency and experiment

Anatomical region	**Co-ordinates**			**AV inc>con**	**TV con>inc**	**Interaction [AV inc>con] x [TV con>inc]**
	x	**y**	**z**			
L lateral inferior temp.	-52	-14	-20	4.9	2.9	5.1
	-52	-6	-32	4.7	1.9	5.0
L antero-medial temp.	-26	-14	-28	3.5	3.4	3.9
				(-26 -14 -30)	(-30 -20 -22)	

Anatomical regions, co-ordinates and Z-scores for interaction between incongruent audiovisual and congruent tactile-visual matching, with corresponding simple main effects.

Figure 8.1: Summary of key findings in left anterior temporal cortex

8.1a: Congruent > incongruent trials in left temporal pole

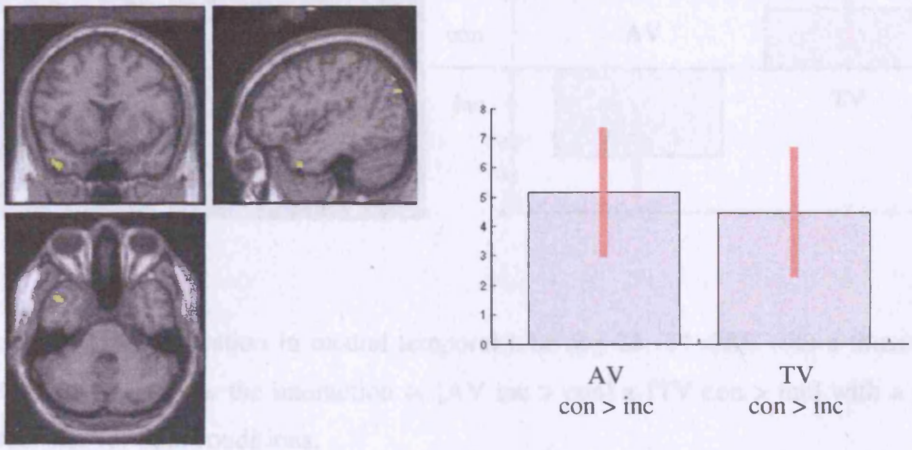


Figure shows temporal pole activation and corresponding effect size for AV and TV congruent > incongruent trials, centred at [-44, 12, -36] and rendered at p<0.001 uncorrected. Activation shown in MNI space on an SPM2 averaged T1 structural MRI image.

8.1b: Incongruent > congruent in left lateral inferior temporal

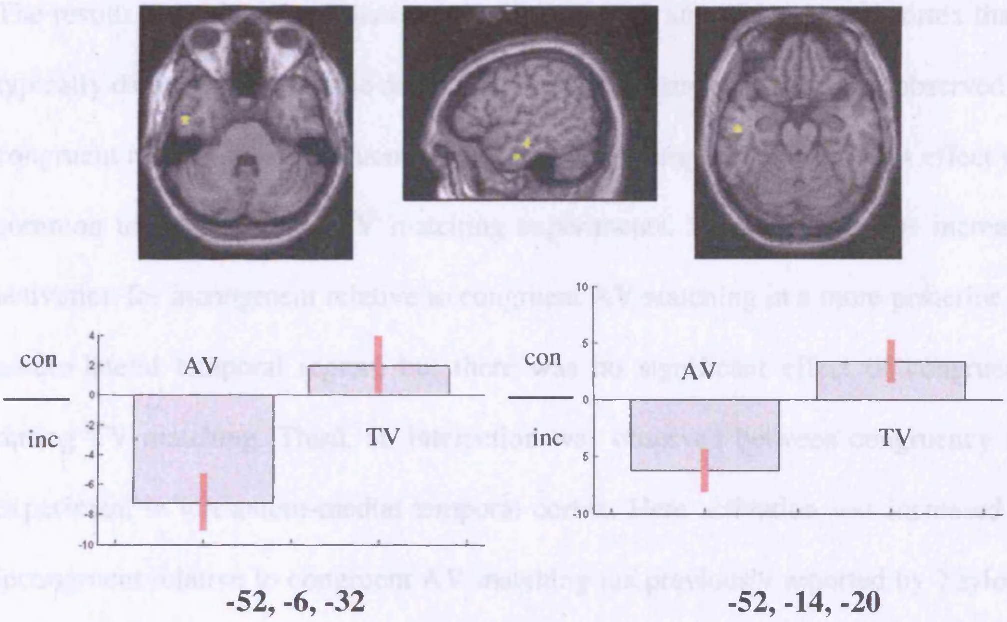


Figure shows activation in two clusters (left section and left plot = [-52, -6, -32], right section and right column = [-52, -14, -20]) of the left infer-lateral temporal cortex, for the effect of matching incongruent > congruent AV pairs. Saggital slice rendered at x = -52, axial slices rendered at z = -32 (left column) and -20 (right column), with a threshold of p<0.001 uncorrected.

8.1c: Interaction of experiment with congruency

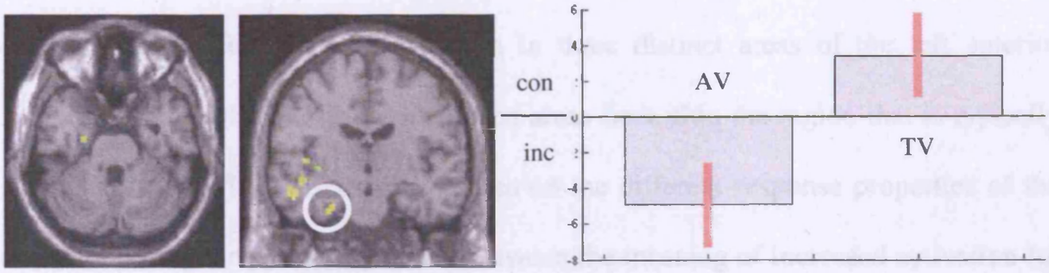


Figure 8.1c shows activation in medial temporal lobe at [-26 -14 -28], with a threshold of p<0.001 uncorrected, for the interaction of [AV inc > con] x [TV con > inc] with a plot of the effect size for both conditions.

Key: AV = audiovisual; TV = tactile-visual; con = congruent, inc = incongruent

8.44 Summary of key findings in left anterior temporal cortex

The results dissociate three distinct effects in the left anterior temporal cortex that is typically damaged in semantic dementia. First, increased activation was observed for congruent relative to incongruent stimuli in the left temporal pole and this effect was common to both AV and TV matching experiments. Second, there was increased activation for incongruent relative to congruent AV matching in a more posterior left antero-lateral temporal region, but there was no significant effect of congruency during TV matching. Third, an interaction was observed between congruency and experiment in left antero-medial temporal cortex. Here activation was increased for incongruent relative to congruent AV matching (as previously reported by Taylor *et al.*, 2006), with the reverse trend for TV matching (as previously reported by Holdstock *et al.*, in submission).

8.5 Discussion

I report here a functional dissociation in three distinct areas of the left anterior temporal lobe. All three of the identified areas lie within the region that is typically damaged in SD. This discussion focuses on the different response properties of the three identified areas, distinguishing between the meaning of increased activation for congruent or incongruent stimuli. This draws on previous functional imaging literature, and then relates these results to the different types of functions that are impaired following anterior temporal lobe damage.

8.51 Increased activation for congruent stimuli in the left temporal pole

Across both experiments, matching congruent relative to incongruent crossmodal pairs increased activation in the left temporal pole. This region has been associated with semantic processing in many other studies. Notably, the peak co-ordinates [-44, 12, -36] are very close to those associated with semantic priming (Crinion *et al.*, 2006; Mummery *et al.*, 1999b; Rossell *et al.*, 2003). For example, the peak co-ordinates associated with semantic priming in Rossell *et al.* (2003) were reported at [-40, 14, -34]. Nevertheless, there are some important differences between the pattern of results in this experiment compared to those reported in semantic priming experiments. In this experiment, left temporal pole activation was higher when subjects matched two crossmodal inputs with the same meaning or shape than when they matched two crossmodal inputs with different meanings or shapes. By contrast, in the semantic priming experiments listed above, the left temporal pole activation was higher when subjects focused on visually presented object names that were preceded by an unrelated object name than a related object name.

This contrasting pattern of activation in the different paradigms demonstrates the context-sensitive nature of left temporal pole responses. The determining factors clearly do not depend on the amount of semantic information that is presented. If this had been the case, then activation would be higher when two semantically different objects are presented than when the same object is presented. Nor can the driving factor relate to the integration of semantic information *per se*. If this had been the case then it would not have been expected to see increased activation for congruent relative to incongruent tactile and visual shapes which do not have semantic associations.

Other functional imaging experiments have associated the left temporal pole with discrimination of unique items such as faces (Damasio *et al.,* 1996; Gorno-Tempini and Price, 2001; Grabowski *et al.,* 2001; Tranel *et al.*, 1997), or subordinate decisions on animals or objects (Gauthier *et al.*, 1997; Rogers *et al.*, 2005). For example when normal subjects make specific (e.g. responding to "Labrador") relative to general (e.g. responding to the same stimulus as "animal") category verification judgements (Rogers *et al.,* 2005, 2006). These observations support the hypothesis that the temporal poles are involved in the relational details associated with an object which are more complex for unique than non-unique items (Damasio *et al.,* 1996; Tranel *et al.,* 1997). Drawing on evidence of the correspondence between functional imaging data from normal subjects and neuropsychological and structural brain imaging data from patients with SD, Rogers *et al.* (2006) suggested that the anterior temporal poles mediate communication between modality-specific perceptual, motor and linguistic regions. In other words, they act as a crossmodal mediational "hub". The finding that activation is higher for congruent than incongruent stimuli supports the theory that the left temporal pole is involved in successful integration of input from distributed regions of modality-specific cortex. Moreover, I extend previous findings by providing evidence that this region is engaged when inputs can be successfully combined into a perceptual or semantic "whole", independent of meaning or the modality of input.

8.52 Activation for incongruent objects in left infero-lateral temporal cortex

In the audiovisual experiment only, increased activation was observed in two clusters of the left infero-lateral temporal cortex for incongruent relative to congruent audiovisual conceptual matching. The lack of significant congruency effect at these locations during the tactile-visual shape matching experiment suggests that the left infero-lateral temporal regions are primarily concerned with conceptual processing or not as responsive to tactile input. The experimental design does not allow me to determine whether these regions respond to meaningful tactile stimuli. However, several previous studies have demonstrated that the left infero-lateral temporal cortex is involved in semantic processing of both auditory and visual inputs (Marinkovic *et al.,* 2003; Scott *et al.*, 2003). For example, a recent study has shown that the implicit comprehension of auditory and visual sentences converges in the left infero-lateral temporal cortex (Spitsyna *et al.,* 2006). The peak infero-lateral co-ordinates reported here [-52, -14, -20] are with 8mm of the peak of a large cluster reported by Spitsyna *et al.* (2006) [-58, -6, -16].

Increased activation when audiovisual inputs refer to two different object concepts (i.e. the incongruent condition) relative to one single object concept (i.e. the congruent condition) demonstrates that the function of the left infero-lateral temporal cortex is different from that of the left temporal pole. Whereas the left temporal pole appears to play a role in integrating multi-modal inputs, irrespective of whether they are meaningful or meaningless, the more posterior left infero-lateral temporal regions appear to increase their responses in proportion to the amount of semantic information that is presented (i.e. when two different object concepts are presented relative to one object). Nevertheless, it is likely that multi-modal object processing involves the interactions between the different anterior temporal regions.

In generative models of object processing, perceptual categorisation and representation are described in terms of context dependent backward connections and prediction (Friston and Price, 2001). The role of the backwards connections provides contextual guidance to lower levels through prediction of the lower area's inputs. When this prediction is incomplete or incompatible with the lower area's input, an error is generated that causes changes in the higher area until there is a reconciliation. By this account, it is proposed that activation in the left infero-lateral cortex reflects a mismatch or error between two conflicting semantic inputs. When error and left infero-lateral temporal activation are high, activation in the temporal pole is low. In contrast, when error and left infero-lateral temporal activation are low, activation in the temporal pole is high. This account is consistent with other proposals that suggest a hierarchy of object processing that proceeds from posterior regions (driving forward and bottom-up) to anterior regions (exerting top-down backward connections). Future experiments could test these predictions using functional connectivity analyses (e.g. dynamic causal modelling), to identify the direction of the driving force.

8.53 Activation by crossmodal inputs in antero-medial temporal lobe

A weaker effect of incongruent relative to congruent audiovisual pairs was also observed in the left medial temporal cortex. Again, this suggests sensitivity to the amount of semantic information or the mismatch between two different object inputs. However, activation in the left medial temporal region was also significantly higher for congruent than incongruent tactile-visual inputs. In other words, activation was higher for 1) incongruent stimuli during audiovisual conceptual matching, and 2) congruent stimuli during the tactile-visual perceptual matching. Although the size of the effects and their interaction were relatively small ($p<0.001$ uncorrected for multiple comparisons), both effects are consistent with previous reports and predictions. The effect of audiovisual object incongruency in the left medial temporal lobe has previously been reported by Taylor *et al.* (2006) whereas the effect of tactile-visual shape congruency has previously been reported by Holdstock *et al.* (in submission).

Taylor *et al.* (2006) suggested that the increased activation for incongruent audiovisual objects was due to integration of conceptual knowledge. In contrast, the results of Holdstock *et al.* (in submission) suggest that activation reflects the integration of non-semantic perceptual information. This is more consistent with the non-human primate literature, where lesions to the perirhinal and entorhinal cortex have been found to disrupt crossmodal processing when pairs of previously learned visual stimuli are associated (Higuchi and Miyashita, 1996) and during tactile-visual associations (Goulet and Murray, 2001). Interestingly, there is some evidence to suggest that damage to these medial temporal regions in humans can also impair performance on tasks of tactile-visual matching (Shaw *et al.*, 1990; Holdstock, 2005), although contradictory evidence (Nahm *et al.,* 1993) means that this line of evidence requires further testing.

The contrasting effects of congruency during audiovisual and tactile-visual matching suggest that the left medial temporal region plays a different role in the object processing hierarchy of the two tasks. During audiovisual conceptual matching, responses increase when there is a mismatch (error) between two inputs. During tactile-visual perceptual matching, responses increase when the two inputs are matching. These observations are consistent with the notion of dynamic representations and generative models of brain function. Thus, neuronal responses in any given cortical area can represent different things at different times (see Friston and Price, 2001). During tactile-visual perceptual matching, the medial anterior temporal cortex may represent the top of the hierarchy and send top-down backward modulatory inputs to other modality specific regions. In contrast, during audiovisual semantic matching, the medial anterior temporal cortex may provide bottom up forward connections to the temporal pole where predictions are generated and modulated.

The complex pattern of effects observed in the medial temporal region has implications for the relative role of this region in perceptual or mnemonic functions. In the non-human primate literature, there is compelling evidence that perirhinal cortex is involved not only in mnemonic processes during object recognition tasks (Buffalo *et al.*, 1998; Meunier *et al.*, 1993; Zola-Morgan *et al.*, 1989) but is also engaged during perceptual discrimination tasks with no memory component (Bussey *et al.*, 2003; Murray and Gaffan, 1994). At an anatomical level, this region is known to have afferents from many uni-modal regions, making it a likely candidate for the integration of perceptual inputs (Suzuki, 1996). For instance, perirhinal cortex has connections predominantly to the ventral visual processing stream, but also to somatosensory cortex and superior temporal gyrus (Murray and Bussey, 1999), making it well placed for processing both tactile-visual and audiovisual information.

The dissociation of medial and lateral temporal functions also has implications for understanding object processing deficits following anterior temporal lobe damage. Patients with selective damage to the medial temporal cortex have more difficulties recognising living than non-living things (Gainotti, 2000; Gainotti *et al.*, 1995). This has been interpreted as the increased demands placed on the perceptual differentiation of complex objects that share many correlated features, such as animals and fruit (Devlin *et al.*, 2002b; Moss *et al.*, 1998; Tyler *et al.*, 2004). When the antero-lateral regions are also damaged, as seen in SD, there is a more encompassing deficit for conceptual knowledge, and the absence of a perceptually-based living/non-living dissociation (Noppeney *et al.*, 2007b). This leads me to reason that successful identification of audiovisual conceptual knowledge requires the recruitment of both antero-medial and antero-lateral temporal cortex, with more polar regions co-ordinating or mediating activation between these regions in order to make the required response. The involvement of any or all of these regions in the intact brain will depend on the type of processing demands required. When no demands are made on conceptual knowledge, as in the case for matching tactile-visual shapes, matching may proceed without involvement of the infero-lateral temporal cortex.

The dissociation in anterior temporal responses also enables a reconciliation of previous disagreements concerning perceptual and conceptual processing in this region. For example Gauthier *et al.* (1997) suggested that temporal pole activation reflected perceptual differentiation demands, whereas Rogers *et al.* (2006) claimed that temporal pole damage (as seen in SD) does not impair the discrimination of visually similar items that are not semantically related. The results of the study reported in this Chapter show that the temporal pole is involved in both perceptual and conceptual matching. However, perceptual matching of congruent tactile visual shapes increased activation in the medial temporal as well as temporal pole area. Therefore, it may be the case that when the temporal pole is damaged, perceptual discriminations can proceed on the basis of medial temporal activation. This hypothesis could be tested by comparing perceptual discrimination in SD patients that have damage to either the temporal pole alone, or to both the temporal pole and medial temporal regions. The prediction is that visual discrimination would be relatively preserved when only the temporal pole was damaged but impaired when both regions were damaged. By contrast, the left medial anterior temporal lobe may not be able to compensate for loss of conceptual matching following temporal pole damage because it does not have the same conceptual response properties as the temporal pole.

8.6 Summary

By comparing patterns of activation from the perceptual and conceptual crossmodal matching tasks, three different anterior temporal lobe functions have been dissociated. In the left temporal pole, where activation was higher for congruent than incongruent objects and shapes, the response pattern is consistent with previous proposals that this region converges multi-modal inputs in order to successfully discriminate a perceptual or semantic "whole". In the left anterior infero-lateral and temporal cortex, where activation was higher for incongruent than congruent audiovisual conceptual matching, the response pattern suggests sensitivity to the amount of conceptual information or the mismatch between two incongruent conceptual items. In the left medial temporal cortex that showed a contrasting pattern of effects in the two experiments, the response pattern suggests it is involved in the discrimination of a perceptual whole (higher activation for matching tactile visual inputs) as well as signalling a mismatch in conceptual inputs (higher activation for mismatching audiovisual inputs). It is proposed that normal object processing requires the interaction of these three areas and that the medial temporal lobe may be able to sustain perceptual discriminations when the left temporal pole is damaged.

Chapter 9

9 Conclusions

9.1 Summary of thesis

Collectively the work documented in this thesis aimed to provide a richer understanding of the anatomico-functional organisation of object processing and representation, through the unique experimental manipulations and comparisons of different types of crossmodal stimuli. More specifically it sought to determine the anatomical and functional levels at which uni-modal and multi-modal inputs are integrated into a perceptual and/or conceptual representation. Relating the neuroanatomical correlates of these data to our current knowledge from functional imaging of humans, as well as patient data and primate lesion studies, should give clues as to the functional processes underlying activation in these anatomical regions.

In this concluding chapter I will firstly summarise the findings from this series of experiments. Following that, I will examine the information that these data provide on the multiple levels at which crossmodal integration can occur. The final section then reiterates the experimental factors that need to be controlled in future studies of crossmodal integration.

9.2 Experimental findings

Chapters 3, 4, 5 and 6 of this thesis addressed the different level(s) at which auditory and visual object integration occurs. This was carried out through experimental manipulation of 1) stimulus material; i.e. verbal and nonverbal, 2) stimulus modality; i.e. uni-modal or audiovisual crossmodal, 3) task; i.e. naming or object decision, and 4) presentation rate; i.e. simultaneous or sequential.

Chapter 3 used PET to investigate the effect of increasing the level of perceptual cues for both uni-modal and crossmodal object stimuli on object naming. Surprisingly there was no effect in the pSTS – a region consistently associated with audiovisual integration (Beauchamp *et al.,* 2004a, 2004b; Calvert *et al.,* 2000, 2001; Sekiyama *et al.*, 2003). Instead, increased perceptual cues activated shared regions associated with both perceptual (bilateral occipital) and conceptual (right anterior fusiform) processing independent of whether the increased cues were within modality (form with colour) or across modality (form with sound). Interestingly the behavioural data complemented the imaging data, with no difference between naming latencies across these two conditions. Moreover, naming latencies were attenuated for increased visual information only, corresponding to the imaging data which showed a decrease in activation (i.e. adaptation) in the antero-medial temporal lobe.

A question remained from Chapter 3 as to whether subjects attended to both crossmodal inputs when performing the task in the scanner. This possible confound was tackled in the fMRI study reported in Chapter 4 by changing the task from naming to object decision. Moreover the design addressed the issue of presenting equivalent stimulus input across all crossmodal and intra-modal trials – considered in this thesis to be a critical confound in previous studies reporting enhanced and super-additive effects for crossmodal relative to uni-modal processing. When both stimulus input and attention to all experimental conditions was controlled, I observed equivalent activation across all three conditions (intra-modal auditory, intra-modal visual and crossmodal audiovisual). Crucially, all conditions activated pSTS, even when activation from the sensori-motor baseline was subtracted. The remarkable consistency of the amodal network activated across all conditions led to the theoretical position that rather than being selective for crossmodal integration, pSTS is involved when associating two meaningful stimuli independent of input modality.

When presenting two auditory stimuli in the intra-modal auditory matching condition stimulus presentation had to be sequential. This was in order that the two auditory stimuli would not create interference at either a perceptual or attentional level (Jancke and Shah, 2002; Lipschutz *et al.*, 2002). The next logical step then, was to investigate whether onset asynchrony was a modulating factor for the involvement of pSTS in crossmodal integration. Using the same stimuli as in Chapter 4, Chapter 5 presented only audiovisual trials and did so with simultaneous auditory and visual stimulus onsets within each trial. A comparison between both sets of audiovisual data from Chapters 4 and 5 confirmed that the same network of activation was consistent across both sequential and simultaneous matching. However, what was revealed was an effect of congruency. Indeed, this effect was in the opposite direction from that predicted on the basis of previous studies using linguistic stimuli: Increased activation was observed in the pSTS region of interest for *incongruent* relative to *congruent* trial pairs. It was concluded that this activation was due to the number of objects simultaneously attended to (two objects for incongruent trials relative to one for congruent trials). This also fits with the data from Chapter 3: When naming objects subjects may have only attended to the visual modality in order to perform the task, therefore the demands on pSTS activation were low.

The fourth issue specific to audiovisual integration investigated in this thesis was the differences that may be elicited through the use of verbal versus nonverbal object information. The design of the study reported in Chapter 5 allowed a parametric investigation of the effect that the level of verbal (or nonverbal) material had on activation within the distributed network identified in Chapters 4 and 5. In line with the prediction that purely verbal matching would increase demands on regions known to be involved in phonological matching, increased activation was observed in the left STS – a region previously associated with amodal verbal relative to nonverbal conceptual processing (Thierry and Price, 2006). Notably, this was more anterior to the pSTS region of interest for audiovisual integration. The contrasting effect for the nonverbal matching condition was less robust. Rather than the effect being observed in the right lateral fusiform gyrus, as would have been expected based on previous comparisons of nonverbal relative to verbal conceptual processing (e.g. Thierry and Price, 2006), activation was more medial and in fact corresponded to a region previously associated with semantic relevance (Mechelli *et al.,* 2006), bottom-up structural object processing (Noppeney *et al.,* 2006) and stored structural knowledge of objects (Chao *et al.,* 1999). Unfortunately the experimental design did not allow further differentiation of the functional process underlying this right medial fusiform activation.

To complement and extend the study of crossmodal integration in this thesis, Chapter 7 investigated tactile-visual integration of abstract shapes. By applying the same experimental presentation parameters to these abstract stimuli as that used for audiovisual pairs, it was also possible to contrast the effects from these different types of integration, as reported in Chapter 8. Taken together, the data presented in Chapters 7 and 8 for tactile-visual matching critically drew together a discrepancy in the previous literature between the findings from human versus non-human primate studies. The effect in the insula for matching crossmodal pairs replicated previous findings from human studies (Banati *et al.,* 2000; Hadjikhani and Roland, 1998). In Chapter 8, I focused on the activation pattern in the anterior temporal lobe. This focus enabled me to investigate the different response properties of the antero-lateral, medial and polar temporal regions. These findings supported the idea that successful recognition of perceptual and conceptual inputs is mediated by the temporal pole, which interacts with medial and antero-lateral temporal regions depending on the demands of the task (perceptual versus conceptual, congruent versus incongruent). The findings additionally provided insight into the differential functions underlying deficits observed in patients with semantic dementia compared with herpes simplex virus encephalitis.

Historically, the study of multisensory integration provided knowledge based on the neuroanatomy of non-human primates, and suggested mechanisms by which inputs from different sensory modalities converge and combine to produce a multimodal percept. With the advent of functional neuroimaging methods, the paradigms used to investigate the anatomico-functional relationship reflected and expanded these primate studies, applying the same principles of multisensory integration to inform our knowledge of how and where the human brain converges and combines different sensory inputs into a unified percept. The studies in this thesis have questioned the validity of using such an approach when investigating both the higher level integration of meaningful multimodal inputs and the perceptual representations of abstract crossmodal inputs. These questions have been driven by the broad issues of: 1) the level at which crossmodal integration is taking place (sensory versus perceptual versus conceptual), and 2) the effect of certain experimental factors that modulate neuronal responses. The next two sections will deal with these questions in turn.

9.3 Crossmodal processing at multiple levels

It is vital to consider the multiple levels at which crossmodal integration is taking place. Early primate studies used low-level sensory stimuli to define regions involved in audiovisual integration. Although useful, the question has to be whether application of knowledge from these early studies is relevant to those reported in this thesis and those published in the current human literature. This is because the functional level at which integration can take place will depend on both the type of stimuli and the action required on it. For example, integrating meaningless audio tones and visual spheres is clearly a very different functional process from assessing whether a previously encountered auditory environmental sound matches a visual picture of a nameable concept.

Despite these stimulus differences, the pSTS has been a consistent region of interest across multiple different functional tasks. For meaningless sensory information, neuronal responses combine low-level information that subjects have been trained to recognise specific to that input. In contrast, the many different types of knowledge associated with a stored semantic representation such as an animal is encoded and expanded upon over many years of experience. Does access to this information depend upon the modality in which it is presented? A large degree of literature has demonstrated that conceptual processing regions are amodal in nature, activating the same neuronal regions independent of input modality (Bookheimer *et al.*, 1998; Booth *et al.*, 2002; Noppeney *et al.*, 2005, 2006; Spitsyna *et al.*, 2006; Thierry and Price, 2006; Vandenberghe *et al.*, 1996). The question then is: Does integration of different sensory input modalities occur within this amodal system at a semantic level, or does integration of crossmodal signals occur at a sensory or perceptual level prior to accessing stored knowledge?

The data in this thesis demonstrate that integration occurs both within an amodal conceptual system and at a perceptual level prior to accessing an amodal representation, with differential neuronal responses depending upon whether subjects are naming or matching. For example, the PET study of naming reported in Chapter 3 found that increased perceptual input increased activation in regions involved in both perceptual (bilateral occipital) and conceptual (anterior fusiform) processing. Moreover, activation in the region of interest in antero-medial temporal lobe showed an increased trend for audiovisual inputs and a corresponding decrease for appropriately coloured objects. When subjects made conceptual object matching decisions on crossmodal inputs, as tested in Chapters 4 and 5, activation was observed across the conceptual network. This was shown for all types of matching (intra-modal auditory, intra-modal visual and crossmodal audiovisual), even when sensori-motor processing was controlled. Chapter 6 reported differential effects within the conceptual system, demonstrating a modulation of conceptual matching depending on whether inputs were verbal or nonverbal in nature. Interestingly, tactile-visual shape matching did show an effect specific to perceptual crossmodal relative to uni-modal integration in the insula. Combining these data with those from the study of audiovisual integration (reported in Chapter 8) demonstrated dissociations between the different types of abstract perceptual versus meaningful conceptual processing.

In summary, the findings across the group of experiments reported in this thesis showed differential effects of perceptual and conceptual level integration depending upon the task (naming versus matching decision), the modality (tactile-visual versus audiovisual) and the stimulus material (verbal versus nonverbal, perceptual versus conceptual). There were two main brain regions of interest: the pSTS and the anterior temporal lobes. I will now look at the involvement of these anatomical regions in more detail.

9.31 Posterior superior temporal sulcus

One of the principles applied to identifying regions where inputs are actively integrated is the enhanced or super-additive neuronal response for crossmodal relative to uni-modal inputs. In anatomical terms, this response enhancement has been observed sub-cortically in the superior colliculus and cortically in a number of temporal lobe regions, but predominantly in the posterior superior temporal sulcus. It is interesting to note that, although one of the initial studies in humans did not find an enhanced or super-additive response in pSTS (Calvert *et al.,* 1997, 1999), the search to designate this region as a site of integration continued. Indeed, as discussed throughout the preceding chapters, an ever increasing number of studies now report activation in this region for a range of crossmodal audiovisual stimuli, including meaningless tones & shapes, linguistic stimuli, and higher-level meaning-driven (conceptual) integration. However, the findings from the studies reported in this thesis suggest that these conclusions need to be moderated and contextualised.

When subjects were required to name meaningful objects, there was no increased activation in pSTS for audiovisual relative to uni-modal inputs (Chapter 3). As shown in Chapter 4, pSTS was equivalently activated independent of whether subjects were matching two meaningful inputs within modality or across modality. Interestingly though, Chapter 5 found that pSTS was more activated by crossmodal inputs when non-matching stimuli were presented simultaneously rather than sequentially. What bearing does this have then on the role for pSTS in integration? The fact that there was a differential effect for incongruent meaningful inputs demonstrates that integration (at a conceptual level) is modulated top down by prior knowledge and not by the modality of bottom-up perceptual input *per se*. Although this does not rule out the involvement of pSTS in the integration, or binding, of low-level audiovisual inputs such as meaningless tones and spheres, the summary of findings provided in Chapter 4 demonstrates that pSTS involvement does not extend to linguistic or conceptual integration when experimental factors such as attention, meaning and level of inputs are controlled. The key message here is that, at a conceptual level of integration, activation is not dependent on input modality, but is dependent on factors related to the meaningful nature of the stimuli and the way in which integration is defined and investigated.

9.32 Anterior temporal lobe

The findings across my studies provide insights into the function of the anterior temporal lobes thereby consolidating a seemingly conflicting group of findings from the normal, clinical and non-human primate literature. In Chapter 3, a region in the antero-medial temporal lobe (previously associated with category selective responses to natural kinds) showed differential responses depending upon the type of perceptual inputs. Specifically, this effect arose from decreased activation for increased uni-modal perceptual cues (form with colour), with a corresponding trend for increased activation in response to audiovisual inputs. These effects corresponded to the naming latencies, with faster responses to increased visual cues compared with increased audiovisual information. It was concluded that this antero-medial temporal region was driven by naming at an amodal conceptual rather than perceptual level, a region strongly linked to semantic processing from studies of patients with semantic dementia and herpes simplex virus encephalitis (Barbarotto *et al.*, 1996; Brambati *et al.*, 2006; Davies *et al.*, 2004; Kapur *et al.*, 1994; Noppeney *et al.*, 2007b).

Chapter 8 specifically investigated the regions involved in different kinds of crossmodal matching and its modulation by the congruency of the stimulus pairs. Three key regions were functionally identified: antero-medial temporal, temporal pole and anterior infero-lateral temporal cortex. Firstly, the conceptual properties of the medial temporal cortex were supported when it was found that the medial temporal cortex was more activated by matching incongruent than congruent audiovisual object pairs that could only be matched at a conceptual level. These data support the view that the medial temporal lobes, including perirhinal cortex, contribute to object recognition (Murray and Bussey, 1999; Murray and Richmond, 2001; Taylor *et al.*, 2006). However, this congruency effect was found to interact with stimulus type. While audiovisual activation was higher for incongruent than congruent pairs, tactile-visual activation was higher for congruent than incongruent pairs. Mnemonic or object recognition processing can not explain the effect of congruency in the tactile-visual experiment because the stimuli were unfamiliar shapes that could only be matched at the perceptual level. The congruency by modality interaction therefore highlights the conflicting arguments regarding the nature of processing (perceptual versus mnemonic) in the medial temporal lobes (Buckley & Gaffan, 2006; Bussey & Saksida, 2005; Bussey *et al.,* 2003; Devlin and Price, 2007; Lee *et al.,* 2005, 2006; Murray and Bussey, 1999; Murray and Richmond, 2001; Murray *et al.,* 2005) by showing that both perceptual and mnemonic (conceptual) processing modulate activation in this region depending upon the type of input.

Turning now to the left temporal pole, it was fascinating to see that this area was equivalently activated by congruent relative to incongruent pairs, independent of the type of crossmodal matching (audiovisual or tactile-visual) and the conceptual or perceptual levels at which the inputs needed to be matched. This particular finding is interesting because it brings together two opposing positions as to the role of this region in perceptually- or conceptually-driven processes. For example, as discussed in Chapter 8, Gauthier *et al.* (1997) prescribed a visual perceptual role for this region when demands are high on differentiation between visually similar objects. An alternative position based on behavioural data from patients with semantic dementia is that this region is engaged by conceptually but not perceptually demanding tasks (e.g. Rogers *et al.,* 2006). Although it is difficult to truly separate conceptual versus perceptual processing (Patterson, 2007), the fact that this region is engaged by both conceptual audiovisual matching and perceptual tactile-visual matching, independent of meaning or modality, suggests that the region responds to both perceptually and conceptually driven tasks in the normal population. When neuronal damage in the temporal pole occurs in disease such as semantic dementia, perceptual discrimination which remains relatively intact behaviourally may be carried out by more (relatively less damaged) medial temporal regions.

The third anterior temporal region that was highlighted in Chapter 8 was the anterior infero-lateral temporal cortex which was more activated by incongruent than congruent meaningful audiovisual pairs, with no effect in the perceptual tactile-visual task. Importantly, this reflects data demonstrating conceptually-mediated processing in semantic dementia patients and functional imaging studies of normal processing, with increased conceptual input increasing neuronal responses (Noppeney *et al.*, 2007b; Rogers *et al.*, 2006).

I concluded from Chapter 8 that these three anterior temporal regions act in concert depending upon the demands placed on functional object processing. From the perspective of generative models of object processing (Friston and Price, 2001), the common neuronal regions (for instance the medial anterior temporal region that responded to both audiovisual mismatches and a tactile-visual match) are viewed as context-sensitive, dynamic responses. During a task which engages conceptually-mediated processing, the lateral and medial anterior temporal regions are more heavily engaged when a mismatch occurs between inputs. During perceptually driven (abstract) integration, the antero-medial region is more responsive to matching than non-matching inputs. On the other hand, the temporal pole responded to both perceptual and conceptual congruent matching, suggesting that this region acts as a modality- and material-independent hub mediating the successful convergence of perceptual and conceptual inputs.

Based on the data presented here, and that of many other studies reporting medial temporal activation for both perceptual and conceptual tasks, the anatomical connectivity of medial temporal lobe structures (perirhinal, entorhinal and hippocampal cortices) with a range of regions (including polar and infero-lateral temporal cortex, visual and auditory association cortices and somatosensory association areas of the insula) predicts functional involvement of this region in the convergence and integration of multiple modalities of stimulus processing. The interactions shown here with additional lateral and polar temporal regions make a strong case for investigating the temporal characteristics of these responses. This may provide a better understanding of the regions engaged by bottom-up, driving inputs and those involved in the higher level modulatory top-down influence.

9.33 Multiple regions, processes and levels

The underlying aim of the anatomico-functional data reported in this thesis has been to shift the focus of investigation away from responses to crossmodal inputs of selective brain regions, such as the pSTS and insula, and reframe the investigation of crossmodal object processing into an understanding of the way that multiple regions interact depending upon factors such as task and stimulus differences. It has been demonstrated here that the pSTS is not specific to audiovisual integration, but is engaged by modality-independent conceptual processing. In contrast the temporal pole is engaged by successful crossmodal integration. Importantly however, this is not specific to audiovisual conceptual inputs, responding also to successful integration of tactile-visual inputs. Furthermore, the temporal pole needs to be viewed not in isolation as a region functionally specialised for crossmodal integration, but within the context of interactions with a network of temporal regions. Is the insula then the only region found in this group of studies that is specific for crossmodal integration? Increased activation observed in the insula in response to crossmodal tactile-visual relative to intra-modal percepts is an important replication of previous studies. However, this effect was not specific to congruent tactile-visual processing, being also activated by incongruent pairs. It is not possible then to conclude that this region is selective for successful crossmodal integration, rather it is more plausible that the insula forms part of a network of regions involved in crossmodal perceptual processing.

9.4 Experimental factors

One of the important contributions that this thesis makes is the identification of a number of problematic issues in experimental design. Future paradigms used to investigate crossmodal integration need to take into account the influence of unbalanced stimulus inputs across trial types, as well as attentional modulation brought about by the manipulations of congruency and temporal presentation rate. These issues will now be summarised in turn.

9.41 Level of stimulus input

My studies (in particular Chapters 3,4,5 and 7) have highlighted the need to control the stimulus input across bimodal and uni-modal trials. Studies of multisensory integration have found a super-additive or enhanced effect for crossmodal relative to uni-modal stimuli, yet it has been shown here that these effects are not significant for audiovisual stimuli when level of sensory input is controlled. However in Chapter 7 (tactile-visual integration) an enhanced response was found in the insula for crossmodal relative to intra-modal stimuli, demonstrating that the experimental paradigm (which controlled for stimulus inputs in both audiovisual and tactile-visual experiments) was able to detect regions involved in crossmodal integration. In summary, when stimulus inputs are not controlled an enhanced crossmodal response cannot be distinguished on the basis of early integration processes versus amodal processing at a later stage.

9.42 Congruency

The effect of congruency on crossmodal processing has previously been confounded by experimental factors relating to attention. For example, when task demands are low or when no task is used, subjects may attend to one modality only when crossmodal inputs are incongruent due to the conflict between signals. It was suggested that this was why no increased activation in pSTS was observed in Chapter 3 where naming could proceed on the basis of input from one modality only. Only by ensuring that subjects attend to both modalities of stimulus inputs can the influence of congruency be effectively measured.

The observation that one region (in the antero-medial temporal cortex) showed opposite effects of congruency during conceptual audiovisual matching and perceptual tactile-visual matching, highlights the necessity to view the effects of congruency within the context of additional interconnected mediating regions rather than being seen as a functional specialisation independent of those interactions.

9.43 Synchrony

One of the principles for defining a region specialised for multisensory integration (as developed from early primate studies by Stein and colleagues) was temporal synchrony. An increasing number of functional neuroimaging studies suggest that temporal synchrony is not necessary for observing enhanced responses due to integration at a perceptual or conceptual level. However the confounds due to task and stimulus input differences already highlighted, compound any effects that may be related to processing synchronous versus asynchronous inputs; firstly, asynchronous inputs can reduce attention to one or other of a crossmodal stimulus due to stimulus interference in comprehension, and secondly, presenting synchronous uni-modal (especially auditory) inputs to ensure sensory inputs are equivalent across uni-modal and crossmodal trials is challenging to say the least.

9.5 Final Conclusions

In Chapter 1 the goals of this body of work were described. To recap, the group of experiments presented in this thesis was designed to use different combinations of visual, auditory and tactile stimuli to elucidate the anatomico-functional organisation underlying object processing and representation. In particular, the question was aimed towards determining the different levels at which intra-modal and multi-modal inputs are integrated into a perceptual and/or conceptual representation. There were three main themes that were tackled:

1. Identifying whether integration occurs within an amodal conceptual system or in earlier perceptually-driven regions

2. Identifying crossmodal integration sites over and above those regions engaged by the equivalent uni-modal inputs
3. Identifying common and distinct regional activations for different types of crossmodal integration

To summarise, both perceptual and conceptual regions were engaged when integrating increased sensory input (Chapter 3). However, these common regions were characterised by differential functional interactions with more anterior temporal regions at a later stage in the object processing hierarchy. This modulation of activation in anterior temporal cortex was observed for abstract as well as conceptual inputs (Chapter 8), and was seen to be both task and stimulus dependent. No differential activation was identified for crossmodal relative to uni-modal integration at a conceptual level (Chapters 4 and 5), although different types of material placed greater demands on regions within this system (Chapter 6). For perceptual abstract tactile-visual integration however, the insula was identified as being selectively involved in crossmodal relative to intra-modal perceptual processing (Chapter 7).

What do these set of results mean for future studies? In terms of audiovisual integration it is clear that the level at which integration occurs (e.g. early sensory versus later-stage conceptual) is a vital factor when looking at candidate neuronal regions for integration. Indeed, an increasing literature that provides evidence of modulation in early sensory cortices by a non-preferred input modality suggests that sensory integration does not necessarily engage a specific multisensory region, but may integrate signals in a cooperative way directly, without recourse to a multisensory or supra-modal region (e.g. von Kriegstein *et al.*, 2003). In contrast, at a conceptual level representations are amodal in nature, and hence access to different types of associated information representing the same concept (e.g. the sound it makes, how it looks, where it is typically seen etc.) would be activated independent of the input modality. Effects at the anatomical level arise when differential functional demands are placed on this system, such as with processing verbal versus nonverbal material, but these differences can only be used to infer a functional role when differential demands are not related to the stimulus modality *per se,* i.e. when factors such as attention, congruency or temporal presentation are controlled. Activation was also observed in regions associated with cognitive factors such as semantic relevance (Mechelli *et al.*, 2006). The influence of these kinds of cognitive factors needs to be understood more fully to determine the organising principles of the object processing system.

The study on tactile-visual integration reported provided an important complement to the audiovisual studies. Firstly, I replicated the effect reported in previous studies for the involvement of the insula in crossmodal integration. Because of the relatively small literature using both well-controlled stimuli and appropriate experimental design, this can be considered a robust effect that adds to a body of knowledge on the many functional roles of the insula. Secondly, the combination of tactile-visual and audiovisual crossmodal matching provided novel data on the role of the anterior temporal regions in object processing, drawing together apparently conflicting data from the normal, patient and non-human primate literatures. This area of study now needs to be extended to investigate the integration of meaningful tactile-visual objects to provide a greater understanding of the amodal nature of conceptual knowledge.

Finally, the data presented here have enhanced our current understanding of the different functional processes that may be affected when brain damage occurs, in particular following lesions to the anterior temporal lobes. The influence of perceptual versus conceptual factors in observed behavioural deficits are often difficult to disentangle due to the multi-faceted nature of a patient's object naming and comprehension difficulties. The data here have provided predictions for the relative influence of perceptual and conceptual factors in relation to medial versus lateral damage, as well as providing support for the view that a wider range of stimulus materials and combinations provides a more complete picture of the anatomico-functional organisation of object processing.

10 References

Adams, R. B. & Janata, P. (2002). A comparison of neural circuits underlying auditory and visual object categorization. *Neuroimage.,* 16, 361-377.

Aguirre, G. K., Zarahn, E. & D'Esposito, M. (1998). A critique of the use of the Kolmogorov-Smirnov (KS) statistic for the analysis of BOLD fMRI data. *Mag.Res.Med.,* 39, 500-505.

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

Amedi, A., Jacobson, G., Hendler, T., Malach, R. & Zohary, E. (2002). Convergence of visual and tactile shape processing in the human lateral occipital complex. *Cereb.Cortex,* 12, 1202-1212.

Amedi, A., von Kriegstein, K., van Atteveldt, N. M., Beauchamp, M. S. & Naumer, M. J. (2005). Functional imaging of human crossmodal identification and object recognition. *Exp.Brain Res.,* 166, 559-571.

Andersson, J., Hutton, C., Ashburner, J., Turner, R. & Friston K. J. (2001). Modelling geometric deformations in EPI time series. *NeuroImage,* 13, 903–919.

Banati, R. B., Goerres, G. W., Tjoa, C., Aggleton, J. P. & Grasby, P. (2000). The functional anatomy of visual-tactile integration in man: a study using positron emission tomography. *Neuropsychologia,* 38, 115-124.

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S. & Hyde, J. S. (1992). Time course EPI of human brain function during task activation. *Magn Reson.Med.* 25, 390-397.

Barbarotto, R., Capitani, E. & Laiacona, M. (1996). Naming deficit in herpes simplex encephalitis. *Acta Neurol.Scand.,* 93, 272-280.

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

Beauchamp, M. S., Lee, K. E., Argall, B. D. & Martin, A. (2004b). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron,* 41, 809-823.

Beauchamp, M. S. (2005). Statistical criteria in FMRI studies of multisensory integration. *Neuroinformatics,* 3, 93-113.

Binder, J. R., Rao, S. M., Hammeke, T. A., Frost, J. A., Bandettini, P. A. & Hyde, J.S. (1994). Effects of stimulus rate on signal response during functional magnetic resonance imaging of auditory cortex. *Brain Res.Cogn Brain Res.,* 2, 31-38.

Binder, J. R. (1997). Neuroanatomy of language processing studied with functional MRI. *Clin.Neurosci.,* 4, 87-94.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M. & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *J.Neurosci.,*17, 353-362.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N. & Possing, E. T. (2000). Human Temporal Lobe Activation by Speech and Nonspeech Sounds. *Cereb.Cortex,* 10, 512-518.

Binkofski, F., Dohle, C., Posse, S., Stephan, K. M., Hefter, H., Seitz, R. J. & Freund, H. J. (1998). Human anterior intraparietal area subserves prehension: a combined lesion and functional MRI activation study. *Neurology,* 50, 1253-1259.

Binkofski, F., Buccino, G., Stephan, K. M., Rizzolatti, G., Seitz, R. J. & Freund, H. J. (1999). A parieto-premotor network for object manipulation: evidence from neuroimaging. *Exp.Brain Res.,* 128, 210-213.

Blank, S. C., Scott, S. K., Murphy, K., Warburton, E. & Wise, R. J. (2002). Speech production: Wernicke, Broca and beyond. *Brain,* 125, 1829-1838.

Bodegard, A., Geyer, S., Grefkes, C., Zilles, K. & Roland, P. E. (2001). Hierarchical processing of tactile shape in the human brain. *Neuron,* 31, 317-328.

Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T. A., Gaillard, W. D. & Theodore, W. H. (1995). Regional cerebral blood flow during object naming and word reading. *Hum.Brain.Mapp.*, 3, 93-106.

Bookheimer, S. Y., Zeffiro, T. A., Blaxton, T. A., Gaillard, W. D., Malow, B. & Theodore, W. H. (1998). Regional cerebral blood flow during auditory responsive naming: evidence for cross-modality neural activation. *Neuroreport,* 9, 2409-2413.

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B. & Mesulam, M. M. (2002). Modality independence of word comprehension. *Hum.Brain Mapp.,* 16, 251-261.

Bornkessel, I., Zysset, S., Friederici, A. D., von Cramon, D. Y. & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage.,* 26, 221-233.

Bozeat, S., Lambon Ralph, M. A., Patterson, K., Garrard, P. & Hodges, J. R. (2000). Non-verbal semantic impairment in semantic dementia. *Neuropsychologia,* 38, 1207-1215.

Brambati, S. M., Myers, D., Wilson, A., Rankin, K. P., Allison, S. C., Rosen, H. J., Miller, B. L. & Gorno-Tempini, M. L. (2006). The anatomy of category-specific object naming in neurodegenerative diseases. *J.Cogn.Neurosci.,* 18, 1644-1653.

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K., Zilles, K. & Fink, G. R. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron,* 29, 287-296.

Brett, M., Penny, W. & Kiebel, S. (2004). Introduction to random field theory. In Frackowiak, R. S. J., Friston, K. J., Frith, C. D., Dolan, R. J., Price, C. J., Zeki, S., Ashburner, J. & Penny, W. (Eds.). *Human Brain Function, 2^{nd} Edition.* London: Academic Press.

Bright, P., Moss, H. & Tyler, L. K. (2004). Unitary vs multiple semantics: PET studies of word and picture processing. *Brain Lang,* 89, 417-432.

Bruce, C., Desimone, R. & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J.Neurophysiol.,* 46, 369-384.

Buckley, M. J. & Gaffan, D. (1997). Impairment of visual object-discrimination learning after perirhinal cortex ablation. *Behav.Neurosci.,* 111, 467-475.

Buckley, M. J. & Gaffan, D. (1998). Perirhinal cortex ablation impairs configural learning and paired-associate learning equally. *Neuropsychologia,* 36, 535-546.

Buckley, M. J., Booth, M. C., Rolls, E. T. & Gaffan, D. (2001). Selective perceptual impairments after perirhinal cortex ablation. *J.Neurosci.,* 21, 9824-9836.

Buckley, M. J. (2005). The role of the perirhinal cortex and hippocampus in learning, memory, and perception. *Q.J.Exp.Psychol.B,* 58, 246-268.

Buckley, M. J. & Gaffan, D. (2006). Perirhinal cortical contributions to object perception. *Trends Cogn Sci., 10,* 100-107.

Buffalo, E. A., Reber, P. J. & Squire, L. R. (1998). The human perirhinal cortex and recognition memory. *Hippocampus,* 8, 330-339.

Bushara, K. O., Grafman, J. & Hallett, M. (2001). Neural correlates of auditory-visual stimulus onset asynchrony detection. *J.Neurosci.,* 21, 300-304.

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K. & Hallett, M. (2003). Neural correlates of cross-modal binding. *Nat.Neurosci.,* 6, 190-195.

Bussey, T. J., Saksida, L. M. & Murray, E. A. (2002). Perirhinal cortex resolves feature ambiguity in complex visual discriminations. *Eur.J.Neurosci.,* 15, 365-374.

Bussey, T. J., Saksida, L. M. & Murray, E. A. (2003). Impairments in visual discrimination after perirhinal cortex lesions: testing 'declarative' vs. 'perceptual-mnemonic' views of perirhinal cortex function. *Eur.J.Neurosci.,* 17, 649-660.

Bussey, T. J. & Saksida, L. M. (2005). Object memory and perception in the medial temporal lobe: an alternative approach. *Curr.Opin.Neurobiol.,* 15, 730-737.

Callan, D. E., Callan, A. M., Kroos, C. & Vatikiotis-Bateson, E. (2001). Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. *Brain Res.Cogn Brain Res.,* 10, 349-353.

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

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D. & David, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport,* 10, 2619-2623.

Calvert, G. A., Campbell, R. & 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. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb.Cortex,* 11, 1110-1123.

Calvert, G. A., Hansen, P. C., Iversen, S. D. & 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. & Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *J.Cogn.Neurosci.,* 15, 57-70.

Calvert, G. A. & Lewis, J. W. (2004). Hemodynamic studies of audiovisual interactions. In G. A. Calvert, C. Spence & B. E. Stein (Eds.). *The handbook of multisensory processes*. London: MIT Press.

Capitani, E., Laiacona, M., Mahon, B. & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cogn Neuropsychol., 20*, 213-261.

Caramazza, A., Hillis, A. E., Rapp, B. C., & Romani, C. (1990). The multiple semantics hypothesis: Multiple confusions? *Cogn.Neuropsychol.*, 7, 161-189.

Chao, L. L., Haxby, J. V. & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat.Neurosci.,* 2, 913-919.

Chao, L. L., Weisberg, J. & Martin, A. (2002). Experience-dependent modulation of category-related cortical activity. *Cereb.Cortex,* 12, 545-551.

Chee, M. W., Weekes, B., Lee, K. M., Soon, C. S., Schreiber, A., Hoon, J. J. & Chee, M. (2000). Overlap and dissociation of semantic processing of Chinese characters, English words, and pictures: evidence from fMRI. *Neuroimage.,* 12, 392-403.

Cherry, S. R. & Phelps, M. E. (1995). Positron emission tomography: Methods and instrumentation. In Diagnostic Nuclear Medicine, M. P. Sandler, R. E. Coleman, F. J. T. Wackers, J. A. Patton, A. Gottschalk, and P. B. Hoffer, (Eds.). Baltimore: Williams & Wilkins.

Clarke, S., Bellmann, A., Meuli, R. A., Assal, G. & Steck, A. J. (2000). Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. *Neuropsychologia,* 38, 797-807.

Coslett, H. B. & Saffran, E. M. (1992). Optic aphasia and the right hemisphere: a replication and extension. *Brain Lang,* 43, 148-161.

Cowey, A. & Weiskrantz, L. (1975). Demonstration of cross-modal matching in rhesus monkeys, Macaca mulatta. *Neuropsychologia,* 13, 117-120.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T. et al. (2006). Language control in the bilingual brain. *Science, 312,* 1537-1540.

Damasio, A. R. (1989). Time-locked multiregional retroactivation: a systems-level proposal for the neural substrates of recall and recognition. *Cognition,* 33, 25-62.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature, 380,* 499-505.

Davies, R. R., Graham, K. S., Xuereb, J. H., Williams, G. B. & Hodges, J. R. (2004). The human perirhinal cortex and semantic memory. *Eur.J.Neurosci.,* 20, 2441-2446.

Degerman, A., Rinne, T., Pekkola, J., Autti, T., Jääskeläinen, I. P., Sams, M. & Alho, K. (2007). Human brain activity associated with audiovisual perception and attention. *Neuroimage,* 34, 1683-1691.

Deibert, E., Kraut, M., Kremen, S. & Hart, J.Jr. (1999). Neural pathways in tactile object recognition. *Neurology,* 52, 1413-1417.

Deneve, S & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *J.Physiol.Paris*, 98, 249-258.

Desimone, R. & Gross, C. G. (1979). Visual areas in the temporal cortex of the macaque. *Brain Res.,* 178, 363-380.

Desimone, R. & Ungerleider, L. G. (1986). Multiple visual areas in the caudal superior temporal sulcus of the macaque. *J.Comp.Neurol.,* 248, 164-189.

Devlin, J. T., Moore, C. J., Mummery, C. J., Gorno-Tempini, M. L., Phillips, J. A., Noppeney, U., Frackowiak, R. S., Friston, K. J. & Price, C. J. (2002a). Anatomic constraints on cognitive theories of category specificity. *Neuroimage.,* 15, 675-685.

Devlin, J. T., Russell, R. P., Davis, M. H., Price, C. J., Moss, H. E., Fadili, M. J. & Tyler, L. K. (2002b). Is there an anatomical basis for category-specificity? Semantic memory studies in PET and fMRI. *Neuropsychologia,* 40, 54-75.

Devlin, J. T., Raley, J., Tunbridge, E., Lanary, K., Floyer-Lea, A., Narain, C., Cohen, I., Behrens, T., Jezzard, P., Matthews, P. M. & Moore, D. R. (2003). Functional asymmetry for auditory processing in human primary auditory cortex. *J.Neurosci.,* 23, 11516-11522.

Devlin, J. T., Rushworth, M. F. & Matthews, P. M. (2005). Category-related activation for written words in the posterior fusiform is task specific. *Neuropsychologia*, 43, 69-74.

Devlin, J. T. & Price, C. J. (2007). Perirhinal Contributions to Human Visual Perception. *Curr.Biol.*.17, 1484-1488.

Dick, F., Saygin, A. P., Galati, G., Pitzalis, S., Bentrovato, S., D'Amico, S., Wilson, S., Bates, E. & Pizzamiglio, L. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: evidence from functional magnetic resonance imaging and lesion data. *J.Cogn.Neurosci.,* 19, 799-816.

Downar, J., Crawley, A. P., Mikulis, D. J. & Davis, K. D. (2000). A multimodal cortical network for the detection of changes in the sensory environment. *Nat.Neurosci.,* 3, 277-283.

Dronkers, N. F. (2000). The pursuit of brain-language relationships. *Brain Lang,* 71, 59-61.

Dronkers, N. F., Wilkins, D. P., Van, V. R., Jr., Redfern, B. B. & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition,* 92, 145-177.

Eacott, M. J., Gaffan, D., & Murray, E. A. (1994). Preserved recognition memory for small sets, and impaired stimulus identification for large sets, following rhinal cortex ablations in monkeys. *Eur.J.Neurosci., 6,* 1466-1478.

Easton, R. D., Srinivas, K. & Greene, A. J. (1997). Do vision and haptics share common representations? Implicit and explicit memory within and between modalities. *J.Exp.Psychol.Learn.Mem.Cogn,* 23, 153-163.

Emmorey, K., Grabowski, T., McCullough, S., Damasio, H., Ponto, L., Hichwa, R. & Bellugi, U. (2004). Motor-iconicity of sign language does not alter the neural systems underlying tool and action naming. *Brain. Lang.,* 89, 27-37.

Ernst, M. O. & Bulthoff, H. H. (2004). Merging the senses into a robust percept. *Trends Cogn Sci.,* 8, 162-169.

Ferreira, C. T., Giusiano, B., Ceccaldi, M. & Poncet, M. (1997). Optic aphasia: evidence of the contribution of different neural systems to object and action naming. *Cortex,* 33, 499-513.

Fox, P.T. (1989).. Functional brain mapping with positron emission tomography. *Semin.Neurol.,* 9, 323-329.

Friederici, A. D. & Kotz, S. A. (2003). The brain basis of syntactic processes: functional imaging and lesion studies. *Neuroimage.,* 20, Suppl1, S8-17.

Friederici, A. D., Rüschemeyer, S. A., Hahne, A. & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb.Cortex,* 13, 170-177.

Friedman, D. P., Murray, E. A., O'Neill, J. B. & Mishkin, M. (1986). Cortical connections of the somatosensory fields of the lateral sulcus of macaques: evidence for a corticolimbic pathway for touch. *J.Comp Neurol.,* 252, 323-347.

Friston, K. J., Jezzard, P. J. & Turner, R. (1994). Analysis of functional MRI time-series. *Hum.Brain.Mapp.*, 2, 153-171.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D. & Frackowiak, R. S. J. (1995a). Spatial registration and normalization of images. *Hum.Brain Mapp.,* 3,165-189.

Friston, K. J., Frith, C., Frackowiak, R. S. J. & Turner, R. (1995b). Characterizing dynamic brain responses with fMRI. *NeuroImage,* 2, 166–172.

Friston, K. J., Holmes, A., Worsley, K., Poline, J. B., Frith, C. D. & Frackowiak, R. S. J. (1995c). Statistical parametric maps in functional imaging: A general linear approach. *Hum.Brain Mapp.,* 2, 189-210.

Friston, K. J. & Price, C. J. (2001). Generative models, brain function and neuroimaging. *Scand.J.Psychol., 42,* 167-177.

Gainotti, G., Silveri, M. C., Daniele, A., & Giustolisi, L. (1995). Neuroanatomical correlates of category-specific semantic disorders: a critical survey. *Memory., 3,* 247-264.

Gainotti, G. (2000). What the locus of brain lesion tells us about the nature of the cognitive defect underlying category-specific disorders: a review. *Cortex,* 36, 539-559.

Garrard, P. & Hodges, J. R. (2000). Semantic dementia: clinical, radiological and pathological perspectives. *J.Neurol.,* 247, 409-422.

Gauthier, I., Anderson, A. W., Tarr, M. J., Skudlarski, P. & Gore, J. C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. *Curr.Biol.,* 7, 645-651.

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

Giraud, A. L. & Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *J.Cogn.Neurosci.,* 13, 754-765.

Glaser, W. R. & Glaser, M. O. (1989). Context effects in stroop-like word and picture processing. *J.Exp.Psychol.Gen.,* 118, 13-42.

Gorno-Tempini, M. L. & Price, C. J. (2001). Identification of famous faces and buildings: a functional neuroimaging study of semantically unique items. *Brain,* 124, 2087-2097.

Goulet, S. & Murray, E. A. (2001). Neural substrates of crossmodal association memory in monkeys: the amygdala versus the anterior rhinal cortex. *Behav.Neurosci.,* 115, 271-284.

Grabowski, T. J., Damasio, H., Tranel, D., Ponto, L. L., Hichwa, R. D., & Damasio, A. R. (2001). A role for left temporal pole in the retrieval of words for unique entities. *Hum.Brain Mapp., 13,* 199-212.

Grefkes, C., Weiss, P. H., Zilles, K. & Fink, G. R. (2002). Crossmodal processing of object features in human anterior intraparietal cortex: an fMRI study implies equivalencies between humans and monkeys. *Neuron,* 35, 173-184.

Grill-Spector, K. (2003). The neural basis of object perception. *Curr.Opin. Neurobiol.,* 13, 159-166.

Hadjikhani, N. & Roland, P. E. (1998). Cross-modal transfer of information between the tactile and the visual representations in the human brain: A positron emission tomographic study. *J.Neurosci.,* 18, 1072-1084.

Hampton, R. R. (2005). Monkey perirhinal cortex is critical for visual memory, but not for visual perception: reexamination of the behavioural evidence from monkeys. *Q.J.Exp.Psychol.B,* 58, 283-299.

Hein, G., Doehrmann, O., Muller, N. G., Kaiser, J., Muckli, L. & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J.Neurosci.,* 27, 7881-7887.

Hickok, G., Buchsbaum, B., Humphries, C. & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J.Cogn Neurosci.,* 15, 673-682.

Higuchi, S. & Miyashita, Y. (1996). Formation of mnemonic neuronal responses to visual paired associates in inferotemporal cortex is impaired by perirhinal and entorhinal lesions. *Proc.Natl.Acad.Sci.U.S.A, 93,* 739-743.

Holdstock, J. S., Gutnikov, S. A., Gaffan, D. & Mayes, A. R. (2000). Perceptual and mnemonic matching-to-sample in humans: contributions of the hippocampus, perirhinal and other medial temporal lobe cortices. *Cortex,* 36, 301-322.

Holdstock, J. S. (2005). The role of the human medial temporal lobe in object recognition and object discrimination. *Q.J.Exp.Psychol.B,* 58, 326-339.

Humphrey, G. K., Goodale, M. A., Jakobson, L. S. & Servos, P. (1994). The role of surface information in object recognition: studies of a visual form agnosic and normal subjects. *Perception,* 23, 1457-1481.

Humphreys, G. W., Lamote, C. & Lloyd-Jones, T. J. (1995). An interactive activation approach to object processing: effects of structural similarity, name frequency, and task in normality and pathology. *Memory.,* 3, 535-586.

Humphries, C.,. Willard, K., Buchsbaum, B. & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: an fMRI study. *Neuroreport,* 12, 1749-1752.

Humphries, C., Love, T., Swinney, D. & Hickok, G. (2005). Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. *Hum.Brain Mapp.,* 26, 128-138.

Humphries, C., Binder, J. R., Medler, D. A. & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *J.Cogn.Neurosci.,* 18, 665-679.

Indefrey, P. and Levelt, W. J. M. (2000). The neural correlates of language production. In Gazzaniga, M. (Ed.). *The new cognitive neurosciences*, MIT Press, Cambridge, MA.

Indefrey, P. & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition,* 92, 101-144.

James, T. W., Humphrey, G. K., Gati, J. S., Servos, P., Menon, R. S. & Goodale, M. A. (2002). Haptic study of three-dimensional objects activates extrastriate visual areas. *Neuropsychologia,* 40, 1706-1714.

Jancke, L. & Shah, N. J. (2002). Does dichotic listening probe temporal lobe functions? *Neurology,* 58, 736-743.

Jarvis, M. J. & Ettlinger, G. (1977). Cross-modal recognition in chimpanzees and monkeys. *Neuropsychologia,* 15, 499-506.

Joseph, J. E. & Gathers, A. D. (2003). Effects of structural similarity on neural substrates for object recognition. *Cogn Affect.Behav.Neurosci.*, 3, 1-16.

Kadunce, D. C., Vaughan, J. W., Wallace, M. T., Benedek, G. & Stein, B. E. (1997). Mechanisms of within- and cross-modality suppression in the superior colliculus. *J.Neurophysiol.*, 78, 2834-2847.

Kapur, N., Ellison, D., Parkin, A. J., Hunkin, N. M., Burrows, E., Sampson, S. A. & Morrison, E. A. (1994). Bilateral temporal lobe pathology with sparing of medial temporal lobe structures: lesion profile and pattern of memory disorder. *Neuropsychologia*, 32, 23-38.

Kawashima, R., Watanabe, J., Kato, T., Nakamura, A., Hatano, K., Schormann, T., Sato, K., Fukuda, H., Ito, K. & Zilles, K. (2002). Direction of cross-modal information transfer affects human brain activation: a PET study. *Eur.J.Neurosci.*, 16, 137-144.

Kellenbach, M.L., Brett, M. & Patterson, K. (2003). Actions speak louder than functions: The importance of manipulability and action in tool representation. *J.Cogn.Neurosci*, 15, 30-46.

King, A. J. & Calvert, G. A. (2001). Multisensory integration: perceptual grouping by eye and ear. *Curr.Biol.*, 11, R322-R325.

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M. & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *Neuroimage.*, 37, 1445-1456.

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. *et al.* (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc.Natl.Acad.Sci.U.S.A,* 89, 5675-5679.

Lambon Ralph, M. A., Graham, K. S., Patterson, K. & Hodges, J. R. (1999). Is a picture worth a thousand words? Evidence from concept definitions by patients with semantic dementia. *Brain Lang,* 70, 309-335.

Lambon Ralph, M. A., Powell, J., Howard, D., Whitworth, A. B., Garrard, P. & Hodges, J. R. (2001). Semantic memory is impaired in both dementia with Lewy bodies and dementia of Alzheimer's type: a comparative neuropsychological study and literature review. *J.Neurol.Neurosurg.Psychiatry,* 70, 149-156.

Lee, A. C., Bussey, T. J., Murray, E. A., Saksida, L. M., Epstein, R. A., Kapur, N., Hodges, J. R. & Graham, K. S. (2005a). Perceptual deficits in amnesia: challenging the medial temporal lobe 'mnemonic' view. *Neuropsychologia,* 43, 1-11.

Lee, A. C., Buckley, M. J., Pegman, S. J., Spiers, H., Scahill, V. L., Gaffan, D., Bussey, T. J., Davies, R. R., Kapur, N., Hodges, J. R. & Graham, K. S. (2005b). Specialization in the medial temporal lobe for processing of objects and scenes. *Hippocampus,* 15, 782-797.

Lee, A. C., Barense, M. D. & Graham, K. S. (2005c). The contribution of the human medial temporal lobe to perception: bridging the gap between animal and human studies. *Q.J.Exp.Psychol.B,* 58, 300-325.

Lee, A. C., Bandelow, S., Schwarzbauer, C., Henson, R. N. & Graham, K. S. (2006a). Perirhinal cortex activity during visual object discrimination: an event-related fMRI study. *Neuroimage.,* 33, 362-373.

Lee, A. C., Buckley, M. J., Gaffan, D., Emery, T., Hodges, J. R. & Graham, K. S. (2006b). Differentiating the roles of the hippocampus and perirhinal cortex in processes beyond long-term declarative memory: a double dissociation in dementia. *J.Neurosci.,* 26, 5198-5203.

Leinonen, L., Hyvarinen, J. & Sovijarvi, A. R. (1980). Functional properties of neurons in the temporo-parietal association cortex of awake monkey. *Exp.Brain Res.,* 39, 203-215.

Lerner, Y., Hendler, T., Ben-Bashat, D., Harel, M. & Malach, R. (2001). A hierarchical axis of object processing stages in the human visual cortex. *Cereb.Cortex,* 11, 287-297.

Levy, D. A., Bayley, P. J., & Squire, L. R. (2004). The anatomy of semantic knowledge: medial vs. lateral temporal lobe. *Proc.Natl.Acad.Sci.U.S.A, 101,* 6710-6715.

Levy, D. A., Shrager, Y. & Squire, L. R. (2005). Intact visual discrimination of complex and feature-ambiguous stimuli in the absence of perirhinal cortex. *Learn.Mem.,* 12, 61-66.

Lewis, J. W. & van Essen, D.C. (2000). Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J.Comp Neurol.,* 428, 112-137.

Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R. & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cereb.Cortex,* 14, 1008-1021.

Linden, J. F., Grunewald, A. & Andersen, R. A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area. II. Behavioral modulation. *J.Neurophysiol.,* 82, 343-358.

Lipschutz, B., Kolinsky, R., Damhaut, P., Wikler, D. & Goldman, S. (2002). Attention-dependent changes of activation and connectivity in dichotic listening. *Neuroimage,* 17, 643-56.

Macaluso, E., George, N., Dolan, R., Spence, C. & Driver, J. (2004). Spatial and temporal factors during processing of audiovisual speech: a PET study. *Neuroimage.,* 21, 725-732.

MacSweeney, M., Amaro, E., Calvert, G. A., Campbell, R., David, A. S., McGuire, P., Williams, S. C., Woll, B. & Brammer, M. J. (2000). Silent speechreading in the absence of scanner noise: an event-related fMRI study. *Neuroreport,* 11, 1729-1733.

Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., Pittet, A. & Clarke, S. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage.,* 14, 802-816.

Mapelli, D. & Behrmann, M. (1997). The role of color in object recognition: Evidence from visual agnosia. *Neurocase,* 3, 237-247.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron, 38,* 487-497.

Martin, A., Wiggs, C. L., Ungerleider, L. G. & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature,* 379, 649-652.

Martin, A. (2007). The representation of object concepts in the brain. *Annu.Rev. Psychol.,* 58, 25-45.

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

McRae, K., Cree, G. S., Westmacott, R. & de Silva, V. (1999). Further evidence for feature correlations in semantic memory. *Can.J.Exp.Psychol.,* 53, 360-373.

Mechelli, A., Sartori, G., Orlandi, P. & Price, C. J. (2006). Semantic relevance explains category effects in medial fusiform gyri. *Neuroimage.,* 30, 992-1002.

Meunier, M., Bachevalier, J., Mishkin, M. & Murray, E. A. (1993). Effects on visual recognition of combined and separate ablations of the entorhinal and perirhinal cortex in rhesus monkeys. *J.Neurosci.,* 13, 5418-5432.

Miller, L. M. & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J.Neurosci.,* 25, 5884-5893.

Milner, A. D. & Goodale, M. A. (1993). Visual pathways to perception and action. *Prog.Brain Res.,* 95, 317-337.

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

Moore, C. J. & Price, C. J. (1999). A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain,* 122 (Pt 5), 943-962.

Moss, H. E., Tyler, L.K., Durrant-Peatfield, M., & Bunn, E. (1998). "Two eyes of a see-through": Impaired and intact knowledge in a case of selective deficit for living things. *Neurocase,* 4, 291-301.

Moss, H. E., Rodd, J. M., Stamatakis, E. A., Bright, P. & Tyler, L. K. (2005). Anteromedial temporal cortex supports fine-grained differentiation among objects. *Cereb.Cortex,* 15, 616-627.

Mummery, C. J., Patterson, K., Wise, R. J. S., Vandenbergh, R., Price, C. J. & Hodges, J. R. (1999a). Disrupted temporal lobe connections in semantic dementia. *Brain,* 122, 61-73.

Mummery, C. J., Shallice, T. & Price, C. J. (1999b). Dual-process model in semantic priming: A functional imaging perspective. *Neuroimage,* 9, 516-525.

Mummery, C. J., Patterson, K., Price, C. J., Ashburner, J., Frackowiak, R. S. & Hodges, J. R. (2000). A voxel-based morphometry study of semantic dementia: relationship between temporal lobe atrophy and semantic memory. *Ann.Neurol.,* 47, 36-45.

Munhall, K. G., Gribble, P., Sacco, L. & Ward, M. (1996). Temporal constraints on the McGurk effect. *Percept.Psychophys.,* 58, 351-362.

Murray, E. A. & Mishkin, M. (1985). Amygdalectomy impairs crossmodal association in monkeys. *Science,* 228, 604-606.

Murray, E. A. & Gaffan, D. (1994). Removal of the amygdala plus subjacent cortex disrupts the retention of both intramodal and crossmodal associative memories in monkeys. *Behav.Neurosci., 108,* 494-500.

Murray, E. A. & Bussey, T. J. (1999). Perceptual-mnemonic functions of the perirhinal cortex. *Trends Cogn Sci.,* 3, 142-151.

Murray, E. A. & Richmond, B. J. (2001). Role of perirhinal cortex in object perception, memory, and associations. *Curr.Opin.Neurobiol.,* 11, 188-193.

Murray, E. A., Graham, K. S., & Gaffan, D. (2005). Perirhinal cortex and its neighbours in the medial temporal lobe: contributions to memory and perception. *Q.J.Exp.Psychol.B, 58,* 378-396.

Murray, M. M., Michel, C. M., Grave de, P. R., Ortigue, S., Brunet, D., Gonzalez, A. S. & 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. & 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.

Nahm, F. K., Tranel, D., Damasio, H., & Damasio, A. R. (1993). Cross-modal associations and the human amygdala. *Neuropsychologia, 31,* 727-744.

Narain, C., Scott, S. K., Wise, R. J., Rosen, S., Leff, A., Iversen, S. D. & Matthews, P. M. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cereb.Cortex,* 13, 1362-8.

Noppeney, U., Josephs, O., Kiebel, S., Friston, K. J. & Price, C. J. (2005). Action selectivity in parietal and temporal cortex. *Brain Res.Cogn Brain Res.,* 25, 641-649.

Noppeney, U., Price, C. J., Penny, W. D. & Friston, K. J. (2006). Two distinct neural mechanisms for category-selective responses. *Cereb.Cortex,* 16, 437-445.

Noppeney, U., Josephs, O., Hocking, J., Price, C. J. & Friston, K. J. (2007a). The Effect of Prior Visual Information on Recognition of Speech and Sounds. *Cereb.Cortex,* epub ahead of print.

Noppeney, U., Patterson, K., Tyler, L. K., Moss, H., Stamatakis, E. A., Bright, P., Mummery, C. & Price, C. J. (2007b). Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia. *Brain,* 130, 1138-1147.

Nyberg, L., Habib, R., McIntosh, A. R. & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proc.Natl.Acad.Sci.U.S.A,* 97, 11120-11124.

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

Ogawa, S., Menon, R. S., Tank, D. W., Kim, S. G., Merkle, H., Ellermann, J. M. & Ugurbil, K. (1993). Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *J.Biophys.,* 64, 803-812.

Ojanen, V., Mottonen, R., Pekkola, J., Jaaskelainen, I. P., Joensuu, R., Autti, T. & Sams, M. (2005). Processing of audiovisual speech in Broca's area. *Neuroimage,* 25, 333-338.

Okada, K., Smith, K. R., Humphries, C. & Hickok, G. (2003). Word length modulates neural activity in auditory cortex during covert object naming. *Neuroreport,* 14, 2323-2326.

Olson, I. R., Gatenby, J. C. & 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.

Parker, A. & Gaffan, D. (1998) Lesions of the primate rhinal cortex cause deficits in flavour-visual associative memory. *Behav.Brain Res.,* 93, 99-105.

Patterson, K. (2007). The reign of typicality in semantic memory. *Philos.Trans.R.Soc.Lond B Biol.Sci.,* 362, 813-821.

Paulesu, E., Frith, C.D. & Frackowiak, R.S.J. (1993). The neural correlates of the verbal component of working memory. *Nature,* 25, 342-345.

Pearson, R. C., Brodal, P., Gatter, K. C. & Powell, T. P. (1982). The organization of the connections between the cortex and the claustrum in the monkey. *Brain Res.,* 234, 435-441.

Perani, D., Schnur, T., Tettamanti, M., Gorno-Tempini, M., Cappa, S. F. & Fazio, F. (1999). Word and picture matching: a PET study of semantic category effects. *Neuropsychologia,* 37, 293-306.

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M. & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature,* 331, 585-589.

Phillips, J. A., Noppeney, U., Humphreys, G. W. & Price, C. J. (2002). Can segregation within the semantic system account for category-specific deficits? *Brain,* 125, 2067-2080.

Pietrini, P., Furey, M. L., Ricciardi, E., Gobbini, M. I., Wu, W. H., Cohen, L., Guazzelli, M. & Haxby, J. V. (2004). Beyond sensory images: Object-based representation in the human ventral pathway. *Proc.Natl.Acad.Sci.U.S.A,* 101, 5658-5663.

Prather, S. C., Votaw, J. R. & Sathian, K. (2004). Task-specific recruitment of dorsal and ventral visual areas during tactile perception. *Neuropsychologia,* 42, 1079-1087.

Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorisation and naming. *Q.J.Exp.Psychol.,* 41A, 797-828.

Price, C. J., Wise, R., Ramsay, S., Friston, K., Howard, D., Patterson, K. & Frackowiak, R. (1992). Regional response differences within the human auditory cortex when listening to words. *Neurosci.Lett.,* 146, 179-182.

Price, C. J., Moore. C. J. & Frackowiak, R.S. (1996). The effect of varying stimulus rate and duration on brain activity during reading. *Neuroimage,* 3, 40-52.

Price, C.J., Green, D.W. & von Studnitz, R. (1999). A functional imaging study of translation and language switching. *Brain,* 122, 2221-2235.

Price, C. J., McCrory, E., Noppeney, U., Mechelli, A., Moore, C. J., Biggio, N. & Devlin, J. T. (2006). How reading differs from object naming at the neuronal level. *Neuroimage.,* 29, 643-648.

Raichle, M., Martin, W., Herscovitch, P., Mintun, M. & Markham, J. (1983). Brain blood flow measured with intravenous $H_2{}^{15}O$. *J.Nucl.Med.,* 24, 790-798.

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

Rauschecker, J. P. (1998a). Parallel processing in the auditory cortex of primates. *Audiol.Neurootol.,* 3, 86-103.

Rauschecker, J. P. (1998b). Cortical processing of complex sounds. *Curr.Opin. Neurobiol.,* 8, 516-521.

Rauschecker, J. P. & Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc.Natl.Acad.Sci.U.S.A,* 97, 11800-11806.

Reed, C. L., Shoham, S. & Halgren, E. (2004). Neural substrates of tactile object recognition: an fMRI study. *Hum.Brain Mapp.,* 21, 236-246.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R. & Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. *Psychol.Rev.,* 111, 205-235.

Rogers, T. T., Hocking, J., Mechelli, A., Patterson, K. & Price, C. (2005). Fusiform activation to animals is driven by the process, not the stimulus. *J.Cogn.Neurosci.,* 17, 434-445.

Rogers, T. T., Hocking, J., Noppeney, U., Mechelli, A., Gorno-Tempini, M. L., Patterson, K. & Price, C. J. (2006). Anterior temporal cortex and semantic memory: reconciling findings from neuropsychology and functional imaging. *Cogn Affect.Behav.Neurosci.,* 6, 201-213.

Rossell, S. L., Price, C. J., & Nobre, A. C. (2003). The anatomy and time course of semantic priming investigated by fMRI and ERPs. *Neuropsychologia, 41,* 550-564.

Saffran, E. M., Coslett, H. B. & Keener, M. T. (2003). Differences in word associations to pictures and words. *Neuropsychologia,* 41, 1541-1546.

Saito, D. N., Okada, T., Morita, Y., Yonekura, Y. & Sadato, N. (2003). Tactile-visual cross-modal shape matching: a functional MRI study. *Brain Res.Cogn Brain Res.,* 17, 14-25.

Saito, D. N., Yoshimura, K., Kochiyama, T., Okada, T., Honda, M. & Sadato, N. (2005). Cross-modal binding and activated attentional networks during audio-visual speech integration: a functional MRI study. *Cereb.Cortex,* 15, 1750-1760.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O. V., Lu, S. T. & Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci.Lett.,* 127, 141-145.

Scott, S. K., Blank, C. C., Rosen, S. & Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain,* 123, 2400-2406.

Scott, S. K., Leff, A. P. & Wise, R. J. (2003). Going beyond the information given: a neural system supporting semantic interpretation. *Neuroimage.,* 19, 870-876.

Seifert, L. S. (1997). Activating representations in permanent memory: different benefits for pictures and words. *J.Exp.Psychol.Learn.Mem.Cogn,* 23, 1106-1121.

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

Seltzer, B. & Pandya, D. N. (1978). Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res.,* 149, 1-24.

Sestieri, C., Di Matteo, R., Ferretti, A., Del Gratta, C., Caulo, M., Tartaro, A., Olivetti Belardinelli, M. & Romani, G. L. (2006). "What" versus "where" in the audiovisual domain: an fMRI study. *Neuroimage,* 33, 672-680.

Shaw, C., Kentridge, R. W. & Aggleton, J. P. (1990). Cross-modal matching by amnesic subjects. *Neuropsychologia,* 28, 665-671.

Skipper, J. I., van, W., V, Nusbaum, H. C. & Small, S. L. (2007). Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb.Cortex,* 17, 2387-2399.

Spitsyna, G., Warren, J. E., Scott, S. K., Turkheimer, F. E. & Wise, R. J. (2006). Converging language streams in the human temporal lobe. *J.Neurosci.,* 26, 7328-7336.

Stark, C. E. & Squire, L. R. (2000). Intact visual perceptual discrimination in humans in the absence of perirhinal cortex. *Learn.Mem.,* 7, 273-278.

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

Stein, B. E. & Meredith, M. A. (1993). *Merging of the senses.* Cambridge: MIT Press.

Stein, B. E., Meredith, M. A. & Wallace, M. T. (1993). The visually responsive neuron and beyond: multisensory integration in cat and monkey. *Prog.Brain Res.,* 95, 79-90.

Stoesz, M. R., Zhang, M., Weisser, V. D., Prather, S. C., Mao, H. & Sathian, K. (2003). Neural networks active during tactile form perception: common and differential activity during macrospatial and microspatial tasks. *Int.J.Psychophysiol.,* 50, 41-49.

Stowe, L. A., Paans, A. M., Wijers, A. A., Zwarts, F., Mulder, G. & Vaalburg, W. (1999). Sentence comprehension and word repetition: a positron emission tomography investigation. *Psychophysiology,* 36, 786-801.

Summerfield, Q. (1992). Lipreading and audio-visual speech perception. *Philos.Trans.R.Soc.Lond B Biol.Sci.,* 335, 71-78.

Suzuki, W. A. & Amaral, D. G. (1994a). Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. *J.Comp Neurol.,* 350, 497-533.

Suzuki, W. A. & Amaral, D. G. (1994b). Topographic organization of the reciprocal connections between the monkey entorhinal cortex and the perirhinal and parahippocampal cortices. *J.Neurosci.,* 14, 1856-1877.

Suzuki, W. A. (1996). The anatomy, physiology and functions of the perirhinal cortex. *Curr.Opin.Neurobiol.,* 6, 179-186.

Talsma, D. & Woldorff, M. G. (2005). Selective attention and multisensory integration: multiple phases of effects on the evoked brain activity. *J.Cogn.Neurosci.,*17, 1098-1114.

Talsma, D., Doty, T. J., Strowd, R. & Woldorff, M. G. (2006). Attentional capacity for processing concurrent stimuli is larger across sensory modalities than within a modality. *Psychophysiology,* 43, 541-549.

Talsma, D., Doty, T. J. & Woldorff, M. G. (2007). Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb.Cortex,* 17, 679-690.

Tanaka, Y. Z., Koyama, T. & Mikami, A. (1999). Neurons in the temporal cortex changed their preferred direction of motion dependent on shape. *Neuroreport,* 10, 393-397.

Taylor, K. I., Moss, H. E., Stamatakis, E. A. & Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proc.Natl.Acad.Sci.U.S.A,* 103, 8239-8244.

Thierry, G., Giraud, A. L. & Price, C. J. (2003). Hemispheric dissociation in accessing the human semantic system. *Neuron,* 38, 499-506.

Thierry, G. & Price, C. J. (2006). Dissociating verbal and nonverbal conceptual processing in the human brain. *J.Cogn.Neurosci.,* 18, 1018-1028.

Tranel, D., Damasio, H. & Damasio, A. R. (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia,* 35, 1319-1327.

Tyler, L. K., Stamatakis, E. A., Bright, P., Acres, K., Abdallah, S., Rodd, J. M. & Moss, H. E. (2004). Processing objects at different levels of specificity. *J.Cogn.Neurosci.,* 16, 351-362.

van, Atteveldt, N.M., Formisano, E., Goebel, R. & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron,* 43, 271-282.

van Atteveldt, N. M., Formisano, E., Blomert, L. & Goebel, R. (2007a). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb.Cortex,* 17, 962-974.

van Atteveldt, N. M., Formisano, E., Goebel, R. & Blomert, L. (2007b). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage.,* 36, 1345-1360.

van Hoesen, G. W. & Pandya, D. N. (1975). Some connections of the entorhinal (area 28) and perirhinal (area 35) cortices of the rhesus monkey. III. Efferent connections. *Brain Res.,* 95, 39-59.

van Wassenhove, V., Grant, K. W. & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc.Natl.Acad.Sci.U.S.A,* 102, 1181-1186.

Vandenberghe, R., Price, C., Wise, R., Josephs, O. & Frackowiak, R. S. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature,* 383, 254-256.

Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *J.Cogn.Neurosci.,* 14, 550–560.

von, Kriegstein, K., Eger, E., Kleinschmidt, A. & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Brain Res.Cogn Brain Res.,* 17, 48-55.

von Kriegstein, K. & Giraud, A. L. (2006). Implicit multisensory associations influence voice recognition. *PLoS.Biol.,* 4, e326.

von Kriegstein, K., Kleinschmidt, A. & Giraud, A. L. (2006). Voice recognition and cross-modal responses to familiar speakers' voices in prosopagnosia. *Cereb.Cortex,* 16, 1314-1322.

Vuilleumier, P., Henson, R. N., Driver, J. & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat.Neurosci.,* 5, 491-499.

Warren, J. E., Wise, R. J. & Warren, J. D. (2005). Sounds do-able: auditory-motor transformations and the posterior temporal plane. *Trends Neurosci.,* 28, 636-643.

Warrington, E. K. (1975). The selective impairment of semantic memory. *Q.J.Exp.Psychol.,* 27, 635-657.

Warrington, E. K. & Shallice, T. (1984). Category specific semantic impairments. *Brain,* 107, 829-853.

Warrington, E. K. & McCarthy, R. A. (1994). Multiple meaning systems in the brain: a case for visual semantics. *Neuropsychologia,* 32, 1465-1473.

Wheeler, M. E., Petersen, S. E. & Buckner, R. L. (2000). Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proc.Natl.Acad.Sci.U.S.A,* 97, 11125-11129.

Wise, R. J., Scott, S. K., Blank, S. C., Mummery, C. J., Murphy, K. & Warburton, E. A. (2001). Separate neural subsystems within 'Wernicke's area'. *Brain,* 124, 83-95.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J. & McCarthy, G. (2003). Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb.Cortex,* 13, 1034-1043.

Zola-Morgan, S., Squire, L. R., Amaral, D. G., & Suzuki, W. A. (1989). Lesions of perirhinal and parahippocampal cortex that spare the amygdala and hippocampal formation produce severe memory impairment. *J.Neurosci., 9,* 4355-4370.

Appendices

11 Appendices

Appendix 1

Stimuli used for the experiment reported in Chapter 3, examining the effect of increased perceptual cues on object naming

Manmade objects with a prototypical colour

ambulance
baseballbat
basket
basketball
battery
brick
coin
london bus
fireengine
fireextinguisher
flag
landrover
lego
lipstick
postbox
redwinebottle
ring
rollingpin
taxi
telephonebox
tennisball
trafficcone
trafficlight
woodenspoon

Natural objects with a prototypical colour

apple
banana
bean
broccoli
carrot
coconut
corn
courgette
garlic
leek
lemon
lettuce
melon
mushroom
onion
orange
peach
pear
pepper
pineapple
potato
raspberry
strawberry
tomato

Manmade Objects with a prototypical environmental sound

accordion
bell
bicyclebell
bongos'
camera
can
clarinet
clock
coin
cork
cymbals
dice
door
doorbell
drum
fireengine
flute
gong
guitar
gun
hammer
harp
lawnmower
match
phone
piano
plane
razor
saw
tap
toilet
toothbrush
train
trumpet
whistle
zip

Natural Objects with a prototypical environmental sound

baby (crying)
baby (laughing)
bee
cat
cat
chicken
cow
cricket
crow
cuckoo
dog (barking)
dog (yapping)
donkey
duck
elephant
fly
frog
goat
goose
horse
horse
lion
mosquito
owl
parrot
pig
rattlesnake
rooster
seagull
seal
sheep
snake
sparrow
whale
wolf
woodpigeon

Colours

black
blue
brown
cream
green
grey
orange
pink
purple
red
white
yellow

Appendix 2

Stimuli used for audiovisual integration in both simultaneous and sequential presentation experiments (Chapters 4, 5 and 6)

Manmade objects with a prototypical sound

alarmclock	car	doorbell	hoover	razor	toilet
ambulance	chainsaw	drill	keys	saw	toothbrush
basketball	coin	glass	match	scissors	train
bicyclebell	cork	gun	plane	ship	typewriter
camera	dice	hammer	policecar	tap	whistle
cannon	door	helicopter	racquet	telephone	zip

Natural objects (animals) with a prototypical sound

baby (crying)	cow	duck	horse (neigh)	parrot	sheep
bear	crow	eagle	jaguar	pig	sparrow
bull	cuckoo	elephant	kookaburra	rattlesnake	whale
cat	dog	frog	lion	rooster	wolf
chicken	dolphin	gorilla	mosquito	seagull	woodpecker
chimpanzee	donkey	grasshopper	owl	seal	woodpigeon

Musical instruments

accordion	bongos	doublebass	harp	panpipes	tambourine
bagpipes	castanets	drum (snare)	harpsichord	piano	timpani
banjo	cello	flute	horn	recorder	triangle
bassoon	chimes	gong	maracas	saxophone	trumpet
bell (church)	clarinet	guitar	oboe	sitar	violin
bells (hand)	cymbals	harmonica	organ	sleighbells	xylophone

Appendix 3

Co-ordinates for activation peaks identified in the audiovisual condition reported in Chapter 4, at a threshold of $p<0.05$ corrected for multiple comparisons. The threshold was lowered to $p<0.01$ uncorrected to determine corresponding Z-scores for visual and auditory intra-modal conditions at same (or within 6mm) co-ordinates. Co-ordinates highlighted in bold and marked with an asterisk are the closest peaks to the pSTS ROI.

Anatomical Region	*x*	*y*	*z*	AV	VV	AA
Frontal lobe						
R inf frontal sulcus	44	18	28	Inf	Inf	Inf
"	44	24	28	Inf	8.3	Inf
R inf frontal gyrus	50	24	30	Inf	Inf	Inf
R mid frontal gyrus	40	44	28	5.8	3.6	3.5
"	38	40	30	5.3	2.8	3.6
R lat orbito-frontal	40	54	-14	5.9	3.2	1.8
R ant orbito-frontal	46	48	-16	7.7	4.0	2.4
"	38	50	-16	5.7	2.4	1.6
R insula	34	26	-4	Inf	3.5	6.0
L inf frontal sulcus	-48	14	32	Inf	Inf	Inf
"	-44	16	28	Inf	Inf	Inf
L inf frontal gyrus	-48	26	24	Inf	6.7	7.5
L mid frontal gyrus	-44	44	20	6.7	3.5	3.4
"	-44	52	8	6.3	3.7	3.0
L lat orbito-frontal	-46	50	-2	7.1	4.9	3.8
L inf precentral sulcus	-38	8	32	Inf	Inf	Inf
L insula	-32	24	2	Inf	6.0	Inf
Parietal Lobe						
R postcentral gyrus	62	-18	14	Inf	Inf	Inf
R intraparietal sulcus	36	-52	50	Inf	Inf	6.9
"	30	-70	32	6.2	6.6	30 -70 34 (1.3)
R precuneus	8	-64	50	6.0	2.6	1.5
L postcentral gyrus	-60	-20	32	6.4	4.1	4.4
"	-62	-18	36	6.3	3.1	3.0
"	-60	-20	32	6.4	4.1	4.4
"	-62	-18	36	6.3	3.1	3.0
L intraparietal sulcus	-28	-58	44	Inf	Inf	Inf
"	-30	-52	44	Inf	Inf	7.3
"	-34	-50	46	Inf	Inf	7.6
"	-48	-38	44	7.5	5.5	6.1
Temporal lobe						
R ant transverse gyrus	52	-18	8	Inf	Inf	Inf
R ant sup temp gyrus	54	-4	-2	Inf	Inf	Inf

R ant sup temp gyrus	56	12	-12	Inf	6.4	Inf
"	52	6	-8	Inf	Inf	Inf
R mid sup temp gyrus	60	-14	4	Inf	Inf	Inf
"	66	-26	12	Inf	Inf	Inf
"	60	-28	8	Inf	Inf	Inf
"	56	-26	10	Inf	Inf	Inf
R mid sup temp sulcus	64	-26	-2	Inf	7.5	Inf
"	68	-26	0	Inf	Inf	Inf
"	60	-24	4	Inf	Inf	Inf
R post sup temp sulcus	***56**	**-40**	**8**	**Inf**	**6.0**	**Inf**
R post sup temp gyrus	66	-40	10	Inf	Inf	Inf
R post mid temp gyrus	62	-50	10	Inf	3.8	7.0
L ant sup temp gyrus	-52	-4	-2	Inf	Inf	Inf
"	-56	0	-4	Inf	7.1	Inf
"	-48	-8	-4	Inf	Inf	Inf
"	-54	16	-8	Inf	2.6	5.5
L ant sup temp sulcus	-42	0	-20	5.9	1.8	5.3
L mid sup temp gyrus	-50	-18	6	Inf	Inf	Inf
"	-50	-22	8	Inf	Inf	Inf
"	-50	-30	10	Inf	Inf	Inf
"	-62	-30	12	Inf	Inf	Inf
"	-66	-26	4	Inf	6.9	Inf
"	-62	-34	10	Inf	7.8	Inf
L mid temp gyrus	-68	-32	4	Inf	6.6	Inf
L post sup temp sulcus	***-56**	**-44**	**12**	**Inf**	**Inf**	**Inf**
L post mid temp gyrus	-66	-46	-6	6.8	2.6	4.6
Occipital lobe						
R mid occipital gyrus	22	-102	4	Inf	Inf	Inf
"	28	-98	10	Inf	Inf	Inf
"	30	-98	4	Inf	Inf	7.6
"	36	-88	-6	Inf	Inf	2.7
"	18	-102	10	Inf	Inf	Inf
R cuneus	16	-102	-4	Inf	Inf	Inf
R post fusiform	44	-66	-20	Inf	Inf	2.5
L mid occipital gyrus	-18	-104	2	Inf	Inf	Inf
"	-28	-94	6	Inf	Inf	5.5
"	-28	-98	2	Inf	Inf	5.6
"	-14	-102	10	Inf	7.4	5.9
"	-10	-104	-8	Inf	7.8	7.8
"	-50	-76	-2	Inf	Inf	(VV>AA, z=inf)
L inf occipital gyrus	-38	-84	-16	Inf	Inf	4.9
"	-46	-72	-18	Inf	Inf	3.7
"	-40	-72	-20	Inf	Inf	3.6
"	-48	-62	-20	Inf	Inf	4.5
L lingual gyrus	-20	-94	-12	Inf	Inf	Inf
Cerebellum						
R cerebellum	32	-70	-24	Inf	Inf	7.5
	34	-54	-22	Inf	Inf	4.0

	8	-78	-18	Inf	8.4	7.4
	28	-80	-20	Inf	Inf	7.4
	16	-86	-20	Inf	Inf	5.7
	40	-56	-26	Inf	Inf	3.0
L cerebellum	-34	-54	-24	Inf	Inf	5.5
	-32	-62	-22	Inf	Inf	6.7
	-34	-72	-24	Inf	Inf	7.1
	-28	-78	-22	Inf	Inf	6.4
	-50	-58	-28	Inf	Inf	5.6
	-36	-46	-28	Inf	Inf	4.6
	-2	-80	-20	Inf	6.3	5.8
Thalamus	-10	-18	8	6.0	3.5	4.9

Key

R = right hemisphere, L = left hemisphere, ant = anterior, post = posterior, inf = inferior, mid = middle, sup = superior, lat=lateral, Inf = infinite Z-score.

Appendix 4

Co-ordinates for audiovisual activation peaks reported in Chapter 5 during simultaneous matching (>10 voxels at p<0.05 corrected), with corresponding z-scores for sequential matching at same (or within 6mm) co-ordinates.

Anatomical Region	*x*	*y*	*z*	Z-score	Sequential Z-score
Frontal lobe					
R inf frontal gyrus	50	16	26	6.2	Inf
R inf frontal sulcus	44	14	26	5.8	Inf
R insula	34	24	2	5.8	7.0
R cingulate	2	0	56	Inf	Outside FOV
"	6	14	48	5.7	Outside FOV
R mid frontal gyrus	52	4	50	7.3	Outside FOV
"	42	-2	60	6.0	Outside FOV
R sup frontal gyrus	4	-4	68	6.8	Outside FOV
L inf frontal sulcus	-42	12	26	Inf	Inf
L precentral gyrus	-52	0	40	Inf	-50 4 36, z = Inf
L inf frontal gyrus	-58	10	32	5.3	5.0
L insula	-36	18	0	5.5	6.1
"	-32	24	0	5.4	Inf
L cingulate	-4	0	56	Inf	Outside FOV
"	-2	10	50	Inf	Outside FOV
L precentral gyrus	-36	-6	66	Inf	Outside FOV
"	-38	-4	54	6.9	Outside FOV
Temporal Lobe					
R mid sup temporal gyrus	60	-10	-2	Inf	Inf
"	66	-14	0	Inf	Inf
R mid sup temporal sulcus	58	-24	0	Inf	Inf
R mid sup temporal gyrus	52	-12	2	Inf	Inf
R mid sup temporal gyrus	64	-22	2	Inf	Inf
"	62	-18	6	Inf	Inf
"	58	-28	8	Inf	Inf
R transverse temporal sulcus	50	-18	4	Inf	Inf
R ant sup temporal gyrus	54	6	-8	Inf	Inf
R post sup temporal sulcus	58	-34	4	Inf	Inf
R ant sup temporal sulcus	62	0	-8	Inf	Inf
R planum temporale	52	-26	10	Inf	Inf
R transverse temporal gyrus	44	-26	12	Inf	Inf
L sup temporal gyrus	-48	-18	2	Inf	Inf
L transverse temporal sulcus	-52	-20	4	Inf	Inf
L planum temporale	-42	-30	8	Inf	Inf
L post sup temporal sulcus	-58	-34	10	Inf	Inf
"	-62	-26	6	Inf	Inf
L mid sup temporal gyrus	-62	-16	6	Inf	Inf

"	-56	-12	2	Inf	Inf
L post sup mid temporal gyrus	-64	-44	8	Inf	Inf
L post sup temporal sulcus	-58	-40	10	Inf	Inf
"	-50	-38	12	Inf	Inf
L post sup temporal gyrus	-54	-40	16	Inf	Inf
L ant transverse temporal gyrus	-50	0	-2	7.3	Inf
L ant sup temporal gyrus	-52	14	-14	6.5	7.8
"	-54	12	-10	6.3	Inf
"	-56	10	-4	5.9	7.7
L ant sup temporal sulcus	-58	-6	-6	Inf	Inf
Parietal Lobe					
R intraparietal sulcus	34	-50	46	6.2	Inf
"	32	-60	48	5.7	7.8
R postcentral gyrus	56	-24	54	6.8	Outside FOV
L intraparietal sulcus	-28	-64	46	Inf	Inf
"	-30	-56	40	Inf	Inf
L central sulcus	-40	-34	42	6.7	-40-34 42, z=4.8
L postcentral gyrus	-46	-40	48	6.4	-46 -40 44, z=6.3
L postcentral gyrus	-52	-32	56	7.3	Outside FOV
L postcentral sulcus	-40	-46	62	7.1	Outside FOV
L sup parietal gyrus	-34	-60	58	6.0	Outside FOV
Occipital Lobe					
R cuneus	20	-100	0	Inf	Inf
R calcarine sulcus	20	-90	-8	7.0	Inf
R mid occipital gyrus	38	-88	-16	7.2	38 -86 -16, z=7.8
"	42	-90	-6	6.3	42 -88 -6, z = Inf
"	32	-92	-6	6.2	Inf
"	46	-80	-10	5.9	Inf
L mid occipital gyrus	-32	-96	-4	Inf	Inf
L post fusiform gyrus	-44	-66	-14	Inf	Inf
"	-38	-78	-18	Inf	Inf
"	-48	-62	-20	Inf	Inf
L mid fusiform gyrus	-38	-48	-24	Inf	Inf
"	-42	-48	-14	6.2	6.7
L cuneus	-26	-100	-4	Inf	Inf
L mid occipital gyrus	-12	-104	-4	6.7	Inf
L inf occipital gyrus	-26	-92	-18	Inf	Inf
L gyrus descendens	-4	-100	-10	5.1	-6 -100 -10, z=4.1
Cerebellum					
Right	0	-66	-16	Inf	2 -66 -16, z=3.4
	32	-62	-28	Inf	7.7
	12	-84	-22	7.7	12 -84 -20, z=Inf
	10	-78	-22	7.2	Inf
	48	-62	-26	6.9	Inf
	2	-76	-18	6.8	Inf
	42	-48	-34	6.3	42 -48 -32, z=6.4
	36	-50	-36	6.2	36 -50 -34, z=4.5

	28	-70	-24	6.1	Inf
	24	-78	-22	6.0	Inf
	26	-74	-24	5.8	Inf
	42	-72	-22	6.4	Inf
	38	-46	-38	6.3	38 -46 -34, z=6.1
Left	-44	-68	-24	Inf	-44 -68 -22, z = Inf
	-42	-62	-26	Inf	Inf
	-42	-58	-28	Inf	Inf
	-36	-38	-30	Inf	Inf
	-32	-72	-24	7.7	Inf
	-28	-64	-24	7.3	Inf
	-20	-78	-22	7.3	5.9
	-4	-82	-20	6.5	Inf

Key

R = right hemisphere, L = left hemisphere, ant = anterior, post = posterior, inf = inferior, mid = middle, sup = superior, Inf = infinity Z-score, FOV = field of view of functional data image