

FEATURE ARTICLE

Crossmodal Processing in the Human Brain: Insights from Functional Neuroimaging Studies

Gemma A. Calvert

Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), John Radcliffe Hospital, Oxford OX3 9DU, UK

Modern brain imaging techniques have now made it possible to study the neural sites and mechanisms underlying crossmodal processing in the human brain. This paper reviews positron emission tomography, functional magnetic resonance imaging (fMRI), event-related potential and magnetoencephalographic studies of crossmodal *matching*, the crossmodal *integration* of content and spatial information, and crossmodal *learning*. These investigations are beginning to produce some consistent findings regarding the neuronal networks involved in these distinct crossmodal operations. Increasingly, specific roles are being defined for the superior temporal sulcus, the inferior parietal sulcus, regions of frontal cortex, the insula cortex and claustrum. The precise network of brain areas implicated in any one study, however, seems to be heavily dependent on the experimental paradigms used, the nature of the information being combined and the particular combination of modalities under investigation. The different analytic strategies adopted by different groups may also be a significant factor contributing to the variability in findings. In this paper, we demonstrate the impact of computing intersections, conjunctions and interaction effects on the identification of audiovisual integration sites using existing fMRI data from our own laboratory. This exercise highlights the potential value of using statistical interaction effects to model electrophysiological responses to crossmodal stimuli in order to identify possible sites of multisensory integration in the human brain.

Introduction

Humans, like most other organisms, are equipped with multiple sensory channels through which to experience the environment (Stein and Meredith, 1993). Each sense provides qualitatively distinct subjective impressions of the world. Colour and pitch, for example, have no counterparts in somatosensation, nor is there any equivalent of tickle in audition or vision. Despite the remarkable disparity of these sensations, we are nevertheless able to maintain a coherent and unified perception of our surroundings. These crossmodal capabilities confer considerable behavioural advantages. As well as having the capacity to use this sensory information interchangeably, thus maintaining object recognition skills when deprived of a sense, the ability to combine sensory inputs across modalities can dramatically enhance the detection and discrimination of external stimuli and markedly speed responsiveness (Zahn *et al.*, 1978; Stein *et al.*, 1989; Perrott *et al.*, 1990; Hughes *et al.*, 1994; Frens *et al.*, 1995). Because information from the different senses is typically complementary, the crossmodal integration of sensory inputs often provides information about the environment that is unobtainable from any one sense in isolation (O'Hare, 1991). For example, our subjective experience of taste derives from the conjunction of gustatory and olfactory cues. Given the ubiquitous nature of crossmodal processing for human experience, knowledge of the underlying neurophysiology seems vital to our understanding of human brain function.

Despite considerable psychophysical research on crossmodal

processes, efforts to discern the underlying neurophysiological mechanisms in humans have only really gained momentum over the past decade. Consequently, there is as yet little terminological consensus in this area. To date, a plethora of terms have been used in the context of crossmodal research ('heteromodal', 'multimodal', 'intersensory', 'polysensory', 'multisensory', 'amodal', 'supramodal', 'modality-specific', 'unimodal', etc.). In certain instances, some of these terms have been used interchangeably, despite the fact that in different contexts they can have quite distinctive meanings. For example, the term 'multimodal' has been used to refer both to the presence of multiple sensory inputs, as well as to the use of multiple experimental techniques [e.g. combined magnetoencephalographic (MEG)/functional magnetic resonance imaging (fMRI) research]. This practice of using terms interchangeably across different levels of explanation (e.g. at the behavioural, anatomical and cellular levels) has also led to some common misconceptions. For example, that all tasks involving more than one sensory modality (e.g. crossmodal matching, crossmodal integration) rely on the same underlying mechanisms, and that neuroanatomical multisensory convergence sites necessarily contain cells that synthesize these crossmodal cues. As a review of the literature will show, these assumptions are often unjustified. For the purposes of clarification, in this review, the terms 'unimodal' and 'crossmodal' have been used to refer to behavioural tasks involving one or many senses, respectively. The terms 'sensory-specific' and 'multisensory' have been used to describe the activity of neurons that respond to one sense or more, and the term 'heteromodal' has been used to refer to neuroanatomical areas that receive converging projections from different senses.

To date, our appreciation of the nature and location of neuronal mechanisms underlying crossmodal processing in humans has been inferred largely from animal experiments carried out in several different species, using a plethora of different techniques. Neuroanatomical studies in primates have identified numerous areas where afferents from the different senses converge (Fig. 1). At the cortical level, these heteromodal sites include zones within the superior temporal sulcus (STS), intraparietal sulcus (IPS), parieto-preoccipital cortex, posterior insula and frontal regions including the premotor, prefrontal and anterior cingulate (AC) cortices (Jones and Powell, 1970; Chavis and Pandya, 1976; Seltzer and Pandya, 1978; Seltzer and Pandya, 1980; Mesulam and Mufson, 1982; Pandya and Yeterian, 1985). Anatomical convergence zones have also been found in subcortical structures including the claustrum, the superior colliculus, the supragenulate and medial pulvinar nuclei of the thalamus, and within the amygdaloid complex including rhinal cortex and hippocampus (Turner *et al.*, 1980; Mesulam and Mufson, 1982; Pearson *et al.*, 1982; Fries, 1984; Mufson and Mesulam, 1984).

The convergence of projections from different sensory

systems in specific brain areas does not necessarily imply convergence at the neuronal level. However, electrophysiological studies in a number of species have shown that many of these heteromodal areas do contain cells responsive to stimulation in more than one modality (Benevento *et al.*, 1977; Desimone and Gross, 1979; Bruce *et al.*, 1981; Rizzolatti *et al.*, 1981a,b; Ito, 1982; Vaadia *et al.*, 1986; Baylis *et al.*, 1987; Hikosaka *et al.*, 1988; Duhamel *et al.*, 1991). This provides the potential, at least, for synthesis of different sensory inputs in heteromodal zones. Investigation of the mechanisms by which crossmodal integration is achieved (see below) and the rules that govern this process at the neuronal level have, to date, been confined largely to the superior colliculus – a structure involved in mediating orientation and attentive behaviours (Stein and Meredith, 1993).

Neurons that respond to stimulation in more than one modality have been identified in the deep layers of the superior colliculus of a number of species, including cats (Gordon, 1973; Meredith and Stein, 1983; Peck, 1987), monkeys (Jay and Sparks, 1984) and ferrets (King and Palmer, 1985). Each bi- or trisensory neuron in this structure contains a map of sensory space, one for each sense (visual, auditory, tactile) to which it responds. The different maps overlap each other so that stimuli from different sensory modalities originating in the same spatial location activate the same region of the superior colliculus (Stein *et al.*, 1975; King and Palmer, 1985). Because these maps are also in register with the premotor maps in the superior colliculus (Harris, 1980; McIlwain, 1986), crossmodal coordinate information can be translated directly into an appropriate orientation response [i.e. via saccadic eye movements (Hughes *et al.*, 1994; Frens *et al.*, 1995)]. Many multisensory neurons in the superior colliculus do more than simply respond to stimulation from different senses. Some are capable of transforming the separate sensory inputs into an integrated product – a phenomenon referred to as ‘multisensory integration’. When two or more sensory cues from different modalities appear in close temporal and spatial proximity, the firing rate of these ‘*multisensory integrative*’ (MSI) cells can increase multiplicatively (i.e. beyond that expected by summing the impulses exhibited to each modality in isolation). These crossmodal enhancements are maximal when the individual stimuli are minimally effective – a principle referred to as *inverse effectiveness* (Stein and Meredith, 1993). By contrast, spatially disparate crossmodal cues can produce a profound *response depression*. In this case, a vigorous response to a unimodal stimulus can be substantially lessened or even eliminated by the presence of a spatially incongruent stimulus from another modality (Kadunce *et al.*, 1997). These principles of multisensory integration (sensitivity to temporal and spatial correspondence, response enhancement and depression and the inverse effectiveness rule) identified at the single neuron level in the superior colliculus have also been shown to apply to the orientation and attentive behaviours mediated by this structure (Stein *et al.*, 1988, 1989).

More recently, MSI neurons have been identified in the cortex of cat (Wallace *et al.*, 1992; Wilkinson *et al.*, 1996), rat (Barth *et al.*, 1995) and monkey (Mistlin and Perrett, 1990; Duhamel *et al.*, 1991; Graziano and Gross, 1998). The properties of these MSI cells are similar in many respects to those in the superior colliculus although there are some significant differences in spatial sensitivity that may reflect differences in emphasis on the nature of the information being combined (i.e. spatial versus identity) in collicular and cortical MSI cells respectively (Wallace *et al.*, 1992). The possibility of independent integrative systems in the colliculus and cerebral cortex is supported from studies in cats demonstrating that it is the sensory-specific neurons in

the cortex, rather than their multisensory counterparts, that project to MSI cells in the colliculus (Wallace *et al.*, 1993). Thus, although there is evidence from different experimental approaches suggesting that crossmodal integration in the cerebral cortex may subservise a different role from that in the superior colliculus and is governed by slightly different rules, very little is currently known about the details underlying cortical integrative processes.

In the light of these electrophysiological data demonstrating the presence of MSI cells in heteromodal cortex, it is curious that lesion studies in non-human primates have found little evidence of any involvement of these areas in crossmodal abilities [for reviews see (Ettlinger and Wilson, 1990; Murray *et al.*, 1998)]. Indeed, in reviewing the relevant literature, Ettlinger and Wilson concluded, ‘none of the structures known to receive converging input from more than one sensory system has been shown to be specifically crucial for both the development and display of crossmodal performance’. Instead, they proposed that the apparent synthesis of information from different sensory modalities might be achieved through synchronized activity between sensory-specific cortices. Such cortical synchronization, they argued, could potentially be co-ordinated via a relay station such as the claustrum, which receives and gives rise to projections from the various sensory systems (Pearson *et al.*, 1982). Yet other lesion data suggest a prominent role for rhinal cortex, at least in the development and storage of crossmodal stimulus-stimulus associations (Murray and Mishkin, 1985). To date, the evidence concerning multisensory mechanisms derived from aspiration studies remains inconclusive.

This discrepancy between the electrophysiological and aspiration data may be more apparent than real. Lesion studies have typically investigated the role of heteromodal cortex in crossmodal performance using tasks of crossmodal matching, transfer and recognition (Murray *et al.*, 1998). Single cell recording studies, on the other hand, have focused on neuronal activity during crossmodal integration. Despite an assumption made by Ettlinger and Wilson that all crossmodal behaviours ‘require only one underlying process’ (Ettlinger and Wilson, 1990), there is theoretical and psychological evidence to suggest that different crossmodal tasks may tap rather different crossmodal processes (Stein and Meredith, 1993; Radeau, 1994; Calvert *et al.*, 1998). For example, tasks of *crossmodal matching* involve determining whether previously associated features (or some shared parameter such as shape or size) are matched across *two distinct* objects, whilst tasks of *crossmodal integration* depend on the perception that two or more sensory cues are perceived as emanating from the *same* object (Stein and Meredith, 1993; Radeau, 1994).

A further distinction relates to the nature of the information being compared and/or integrated. Generic features such as intensity, spatial location, rate, rhythmic structure, shape and texture, to name but a few, are examples of *intermodal invariances* (Lewkowicz, 2000). In other words, the information is analogous irrespective of the sensory modality in which it is perceived. The ability to detect equivalence across sensory modalities occurs early in development (Rose and Ruff, 1987; Lewkowicz and Lickliter, 1994) and once learned can be generalized across all relevant combinations of sensory inputs. In contrast, the association between other crossmodal attributes that specify an object must be learned for each prototype. For example, without prior experience, the odour of a banana provides no information concerning its visual appearance. Moreover, once the association between these unimodal features have been learned, they fail to provide any further information

Fig 1

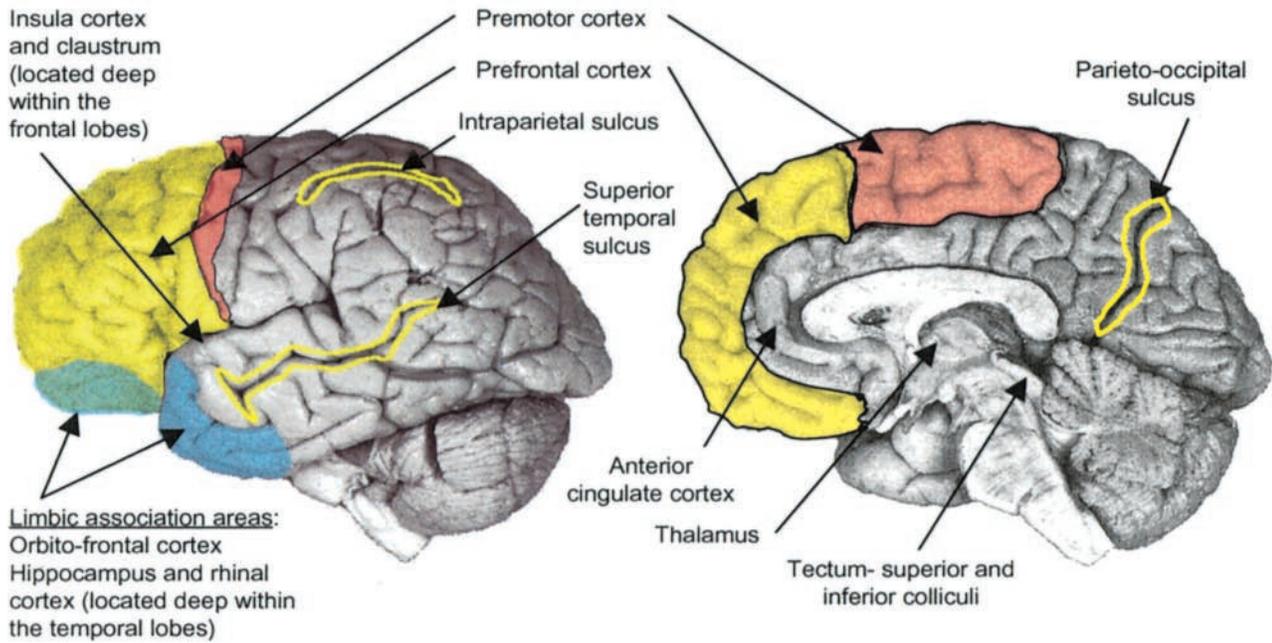


Fig 2

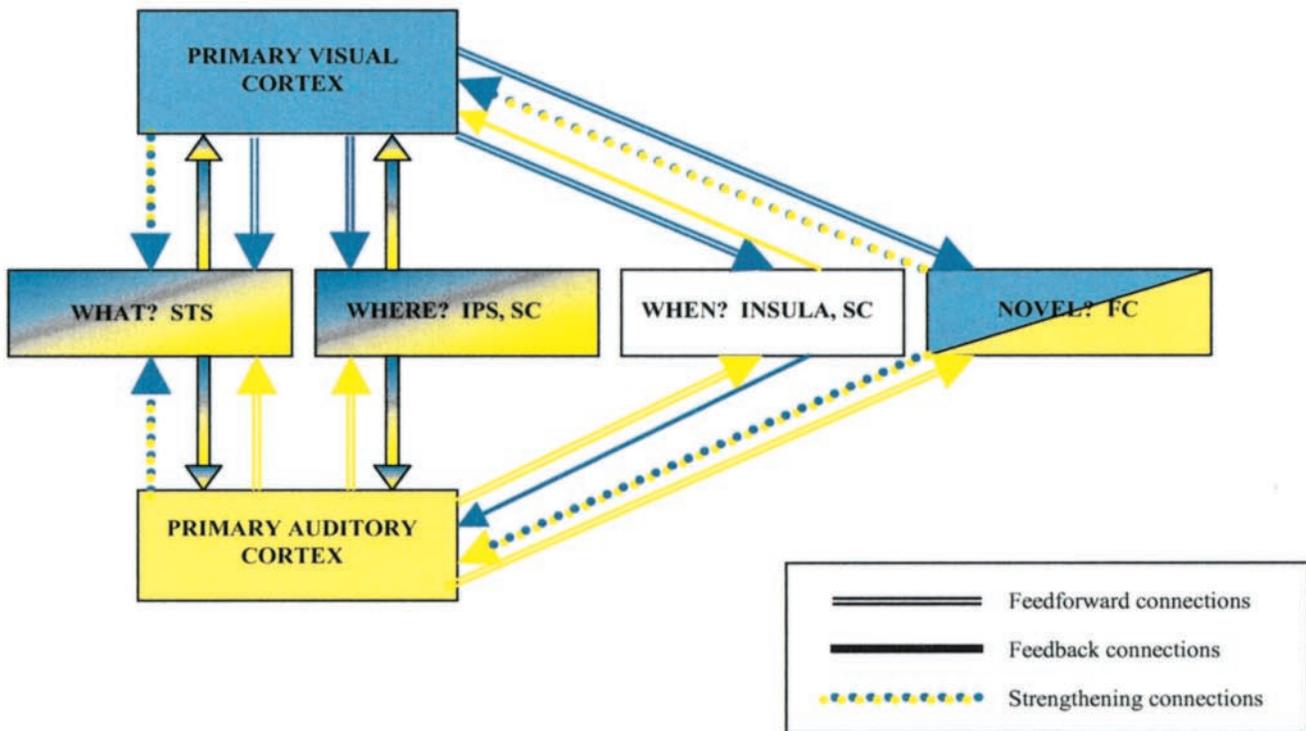


Table 1

Author/s	MODALITIES			TECHNIQUE			PARADIGM				STIMULUS TYPE			ANALYSIS			BRAIN REGIONS						
	AV	VT	AT	PET	fMRI	ERP MEG	MA	ID	S/A	LEA	MS	IA	CR	SU	CO/ SUB	IN	STS	IPS	I-C	FR	SC	UNI	OT
Hadjikhani 98		X		X			X					X			X				X				
Banati 00		X		X			X					X			X		X	X	X	X			X
Calvert 99	X				X			X				X	X		X							X	X
Calvert 00	X				X			X				X	X			X	X			X		X	X
Callan 01	X					X		X				X	X			X	X	X		X		X	X
Rajj 00	X					X		X			X	X	X	X		X	X			X		X	X
Giard 99	X					X		X			X	X				X				X		X	X
Foxe 00			X			X		X				?	?			X						X	X
Bushara 01	X			X				X				X	X		X		X	X	X	X	X	X	X
Calvert 01	X				X			X				X	X			X	X	X	X	X	X	X	X
Macaluso 00		X			X				X			X	X			X		?				X	X
Lewis 00	X				X				X			X	X			X		?	X	X	X		X
Macaluso 00		X		X					X			X	X	X	X				X				
Eimer 99	X					X			X			X							X				?
Bushara 99	X			X					X			X			X	X			X		?		?
Rosenblum 99	X				X					X	X		X									X	
McIntosh 97	X			X							X	X										X	
Gonzalo 00	X				X						X	X								X			X

AV, audiovisual; VT, visuotactile; AT, audiotactile; MA, crossmodal matching; ID, crossmodal identification; S/A, crossmodal localization and crossmodal spatial attention; LEA, crossmodal learning; MS, modality-specific integration; IA, intermodal invariances; CR, meets criteria for binding; SU, superimposition of unimodal tasks; CO, conjunction or subtraction of bimodal versus unimodal/s; IN, interaction effects; STS, superior temporal sulcus; IPS, intraparietal sulcus; I-C, insula-claustrum; FR, frontal cortex; SC, superior colliculus; UNI, sensory-specific cortices; OT, other brain regions.

Colour coding by task: crossmodal matching = red; crossmodal identification of linguistic (blue) and non-linguistic (navy blue) information; crossmodal localisation and spatial attention (pink); crossmodal learning (green).

concerning odour–appearance correlations in other fruits [for a more in-depth discussion of these issues see (Lewkowicz, 2000)].

Whilst electrophysiological studies have predominantly investigated the neural mechanisms underlying the crossmodal integration of spatial information (an example of information that is invariant across senses), lesion studies have typically concentrated on the brain areas involved in learning and retrieving arbitrary crossmodal associations (Gaffan and Harrison, 1991; Murray *et al.*, 1998). These differences in emphasis in both task and informational content between electrophysiological and lesion studies may help to resolve their apparently paradoxical results. Thus, it is worth bearing in mind as we examine the human imaging literature on crossmodal processing that factors such as task and stimuli may have more of an impact on the pattern of brain areas implicated than has been previously suggested (Ettlinger and Wilson, 1990).

As a brief review of the animal literature has illustrated, electrophysiological and aspiration studies can provide only a partial picture of the possible networks involved in perceiving, integrating and responding to multiple sensory inputs. Whilst electrophysiological data have proved invaluable in providing a fine-grained analysis of the response properties of MSI cells in a small number of brain areas, the limitations of this technique make it difficult to study the inputs influencing activity in these cells and the regions subsequently modulated by this activity. Lesion studies, on the other hand, permit several cortical brain

areas to be investigated simultaneously. However, they are unable to determine whether the aspirated areas actually participate in integration or instead contain interconnecting fibres between participant areas. Consequently, lesioning provides a very coarse description of the brain areas that may, or may not, be involved in crossmodal processing, and is limited by the number of putatively different crossmodal tasks that can be examined at any one time. Finally, as the manner in which crossmodal stimuli are processed becomes increasingly sophisticated with evolutionary development (Stein and Meredith, 1993), we must be cautious in extrapolating from studies in non-human animals to humans.

Modern neuroimaging techniques offer an exciting new method of studying these crossmodal interactions in the intact human brain. This paper reviews studies that have used fMRI, positron emission tomography (PET), event-related potentials (ERPs) and electroencephalography (EEG)/MEG to investigate how the human brain copes with multiple sensory inputs. As with any rapidly developing field, there is as yet little consistency either in experimental design or analytic strategy. In an attempt to impose some order on this variability, studies have been divided by task including crossmodal matching, crossmodal integration of identity (crossmodal identification) and spatial coordinate (crossmodal localization) information and crossmodal learning (see Table 1). This method of classification is beginning to implicate certain brain areas in different crossmodal tasks, adding empirical support to the theoretical

Figure 1. Areas of putative heteromodal zones in the human brain shown on lateral (left) and medial (right) views. The identification of these areas has been based on analogy from primate data, human neuropsychological evidence and human functional imaging studies.

Figure 2. A model of crossmodal integration. The diagram presents a tentative model of crossmodal integration based on data from human functional imaging studies. Note that different brain areas appear differentially specialized for the integration of different types of crossmodal information (what, where, when, and for the acquisition of recently learned or arbitrary associations). Two-tone blue–yellow (e.g. for what and where) indicates that the crossmodal information is understood to have been blended; diagonally separated blue–yellow indicates that the crossmodal information may be shared, but not necessarily integrated (e.g. novel associations); white indicates a sensory-specific to sensory-specific relay station where the information passes through but is not bound in these regions. Whilst the frontal cortex may participate in the initial learning of arbitrary crossmodal associations, these associations may later form amodal representations in areas such as the STS as the strength of these associations is enhanced as a function of learning. The possibility of multiple routes of integration may explain discrepancies between lesion and electrophysiological crossmodal data in primates, including for tasks of crossmodal matching (not shown here). Abbreviations: STS, superior temporal sulcus; IPS, inferior parietal sulcus; SC, superior colliculus; FC, frontal cortex.

distinctions discussed above. One potentially avoidable source of variation between putatively similar crossmodal experiments is analytic strategy. The consequences of using of intersections, conjunctions and interactions to identify putative sites of multisensory integration is illustrated using existing fMRI data from audiovisual paradigms acquired in our own laboratory. Suggestions will then be made as to how a consistent approach may be applied to future crossmodal studies to facilitate comparison of results across different combinations of modalities and experimental designs. This review has concentrated on studies examining crossmodal interactions between the auditory, visual and tactile senses. Whilst similar principles may also be operant during the integration of the chemosensory modalities (gustation, olfaction), they will not be discussed in this review due to the current paucity of relevant human imaging studies.

Crossmodal Functional Imaging Studies

Crossmodal Matching

One of the earliest attempts to identify brain areas involved in co-ordinating activity across the senses using functional neuroimaging was a PET study of visuotactile matching (Hadjikhani and Roland, 1988). During scanning, subjects were asked to decide whether pairs of spherical ellipsoids presented visually, tactually or during a crossmodal visuotactile condition were matched on the basis of shape. Conjunction analysis of the crossmodal condition with both unimodal (visual-visual and touch-touch) conditions was designed to uncover brain areas selectively involved in matching shape across the senses. This strategy identified one differential region of activation in the right insula-claustrum (although note that the resolution of PET precludes accurate discrimination of these adjacent structures). On the basis of this finding, the authors found themselves in agreement with Ettliger and Wilson (Ettliger and Wilson, 1990) that there does not appear to be a role for heteromodal cortex in crossmodal matching (at least between visual and tactile information). Instead, crossmodal feature comparisons might be effected via sensory-specific activations, co-ordinated via the claustrum.

In a subsequent PET study of visuotactile matching (Banati *et al.*, 2000), stricter control of the crossmodal and intramodal conditions identified a wider network of brain areas involved in these crossmodal comparisons than had been reported by Hadjikhani and Roland (Hadjikhani and Roland, 1998). In this study, subjects were required to match a metal arc placed on a card obscured from view to one of four circles presented simultaneously on a computer screen. In the unimodal (visual-visual) condition, subjects matched one of four visual circles to a simultaneously visible arc presented in the lower quadrant of the screen. Subtraction of the resulting brain activation maps revealed a network of brain areas, including the inferior parietal lobules, bilateral STS, the AC, left dorsolateral prefrontal cortex (DLPFC), as well as the left insula-claustrum. The authors suggested that one explanation for the discrepancy between these findings and those of Hadjikhani and Roland (Hadjikhani and Roland, 1998) might be the inclusion in the latter study of an additional tactile-tactile matching condition. In contrast to the visuotactile and visual-visual conditions, the tactile-tactile matching control task used by Hadjikhani and Roland required *sequential* rather than *simultaneous* comparison of the two stimuli. To the extent that subjects may have created a visual image of the initially presented tactile stimulus to aid comparison with a subsequently presented tactile object, brain areas involved in crossmodal matching may have been inadvertently

activated in this condition. Thus, subtraction of the tactile-tactile condition from the explicit crossmodal (visuo-tactile) condition would have resulted in an under-appreciation of the areas putatively involved in crossmodal matching.

Clearly, these two experiments are insufficient to determine whether the areas identified are specific to tasks of crossmodal matching or the particular combination of modalities engaged. However, the common finding of insula-claustrum activation provides some preliminary support for the involvement of this region in tasks involving comparison of stimulus features across the senses. Given the somewhat conservative nature of the experimental strategy employed by Hadjikhani and Roland, and the subsequent identification of several other brain areas by Banati and colleagues (Banati *et al.*, 2000), it seems likely that the insula-claustrum forms part of a larger neuronal network, including heteromodal areas, that co-ordinates matching between the visual and tactile modalities. The additional possibility that crossmodal matching may be mediated by two parallel routes, one involving convergence in heteromodal areas and the other direct cross-talk between the sensory-specific cortices mediated via the claustrum, may explain why lesions of heteromodal cortices in other species fail to impair performance on tasks of crossmodal matching and transfer, when the claustrum is intact. Whether or not tasks of crossmodal matching involve the participation of multisensory or MSI neurons, and co-ordinated in the same brain areas as tasks of crossmodal integration, remains to be determined.

Crossmodal Integration

Crossmodal Identification

Crossmodal identification refers to the integration of crossmodal cues relating to the identity of an object (e.g. size, pitch, duration), rather than its position in space. Studies that have investigated where and how the brain synthesizes crossmodal cues relating to feature information can be most usefully divided between those that have used linguistic stimuli (Calvert *et al.*, 1999, 2000; Raji *et al.*, 2001; Callan *et al.*, 2001) and the remainder (Giard and Peronnet, 1999; Foxe *et al.*, 2000; Bushara *et al.*, 2001; Calvert *et al.*, 2001). A further distinction can be made between those studies that have focused on the integration of sensory invariant information (such as temporal onset, duration and/or correlated audiovisual frequency-amplitude information in the case of audiovisual speech) and those that have investigated the crossmodal synthesis of arbitrary crossmodal combinations such as the audible and graphical representation of alphabetic letters, where the acoustic form is arbitrary with respect to the written symbol (Raji *et al.*, 2001).

Studies Involving Linguistic Information. A consistent picture is now beginning to emerge concerning the role of the STS in the synthesis of audiovisual speech. Studies using MEG (Sams *et al.*, 1991) and fMRI (Calvert *et al.*, 1997) had initially demonstrated that visual speech (i.e. lip-reading) was capable of activating areas of auditory cortex previously assumed to be dedicated to processing sound-based signals. In a subsequent fMRI study, Calvert and colleagues showed that when the speaker could be seen as well as heard, activation in the sensory-specific cortices (in this case, auditory and visual) was enhanced compared to the response to either modality alone (Calvert *et al.*, 1999). They argued that increased activity in these modality-specific areas might be the physiological correlate of the subjective improvements in 'hearing' when a speaker's lip and mouth movements are visible (Reisberg, 1987) and superior

'visual localization' of the sound source (i.e. to discriminate who is saying what, in a room full of speakers) when the auditory and visual speech patterns can be matched (Driver, 1996; McDonald *et al.*, 2000). However, none of these studies explicitly investigated the site(s) of convergence between these two sensory modalities.

To address this issue, Calvert and colleagues carried out an fMRI study designed to detect brain areas involved in synthesizing auditory and visual speech signals, using an approach informed by electrophysiological data on multisensory integration (Calvert *et al.*, 2000). Although the response characteristics of MSI cells have only been studied in the context of spatially concordant crossmodal stimuli, behavioural data suggests that similar rules might apply during the crossmodal synthesis of content information. For example, speech comprehension is improved when the speaker can be seen as well as heard (Sumby and Pollack, 1954; Reisberg, 1987), but impaired if the lip and mouth movements are semantically incongruent (Dodd, 1977). In view of these parallels at the neurophysiological and behavioural levels, Calvert and colleagues hypothesized that similar indices of multisensory integration might be detectable at the macroscopic level, manifested in changes in the BOLD (blood oxygen level dependent) response.

To test this hypothesis, subjects were scanned whilst being exposed to varying epochs of semantically concordant or discordant audiovisual speech, to heard speech, seen speech or rest. In the congruent audiovisual condition, a speaker was mouthing the same story as that heard over the headphones. In the incongruent condition, the speaker mouthed a different story. Brain areas involved in the integration of audiovisual speech were then identified on the basis that they exhibited (i) co-responsivity (i.e. were stimulated by each modality in isolation), (ii) a positive interaction effect in response to congruent audiovisual speech [$AV > A + V$] and (iii) a negative interaction effect in the incongruent audiovisual condition [$AV < A + V$]. These criteria are analogous, but not identical to, those imposed at the cellular level to demonstrate multisensory integration. This is because each unit (or voxel) measurable by fMRI reflects the response of a large population of neurons. Anything less than a superadditive interaction could therefore simply reflect the linear summation of responses from two sets of sensory-specific neurons (in this case auditory and visual) rather than multisensory integration responses (which need not be, but often are, superadditive) in MSI cells. The same problem does not arise when recording from an individual neuron where any change in activity can be attributed directly to the cell itself. The requirement for superadditivity (i.e. positive interaction effects) should therefore be seen as a necessary modification of the single cell criterion to ensure validity at the macroscopic level. To extend this statistical approach to response depression (which at the neuronal level is recognized when the bimodal response is less than that elicited by either unimodal cue in isolation, whichever is the greater), negative interactions (the bimodal response is less than the sum of the unimodal responses) were used to represent response suppression at the macroscopic level. Although a less harsh criterion than response depression, negative interaction effects have the advantage of demonstrating that the output has been modified relative to the inputs. Furthermore, as response depression at the neuronal level is observed less frequently than response enhancement, a less harsh criterion may be more appropriate for use in neuroimaging studies where the large number of voxels analysed leads to a serious multiple comparison problem and a reduction in power to detect effects.

The only locus in the brain that met all three criteria was localized in the posterior ventral bank of the left STS [Talairach coordinates (TC): $x = -49$, $y = -50$, $z = 9$]. Other brain areas exhibited positive or negative interaction effects to congruent and incongruent bimodal stimuli, but not both. Most notable were the superadditive effects in auditory and visual cortices in response to semantically congruent audiovisual speech. As these were considerably weaker than those found in the STS, the authors suggested that they reflect multisensory enhancements, rather than integration, replicating earlier findings (Calvert *et al.*, 1999). In view of the fact that not all MSI cells that show positive integration responses necessarily exhibit response depression, particularly in cortex (Wallace *et al.*, 1992), the inclusion of both criteria may be too conservative to detect all the areas involved in the integration process. However, the fact that the STS survived these criteria suggests that it may have a more prominent role in this process, at least for audiovisual speech.

That these interaction effects reflect multisensory integration responses, rather than increases or decreases in attention and/or comprehension, is supported by data from a recent single-sweep EEG study (Callan *et al.*, 2001). In this experiment, audible English words were presented in the presence or absence of noise (multispeaker babble) together with video footage of a speaker mouthing the same words (congruent condition) or non-linguistic facial motion (control condition). According to the principle of inverse effectiveness (Stein and Meredith, 1993), multisensory integration responses are maximal when the unimodal stimuli are least effective. Thus, it was predicted that concordant visual speech should generate a larger response enhancement in MSI cells when the auditory signal is degraded by noise. Alternatively, if the crossmodal interaction effects observed by Calvert *et al.* (Calvert *et al.*, 2000) simply reflected increases or decreases in speech comprehension and/or attention in the presence of congruent and incongruent lip and mouth movements, maximal enhancements should be detected to audiovisual speech in the absence of noise.

Despite predictably better performance when audiovisual speech was presented without noise, the largest interaction effects in the EEG data were detected in the audiovisual speech plus noise condition. Subsequent independent component analysis of the data revealed that these were attributable to two independent components with more energy in the higher-frequency ranges (45–70 Hz) than other conditions. Activity in this frequency band has been previously linked to perceptual binding (Singer and Gray, 1995). The first component was localized to the left superior temporal gyrus, peaking ~150–300 ms after stimulus onset, and was attributed to a stimulus-induced effect. The second component spanned the entire time-course of the signal and was distributed across a network of areas including parietal, occipital, temporal, prefrontal and sensorimotor areas. This was attributed to a task-induced effect. These data are in accordance with those reported by Calvert *et al.* (Calvert *et al.*, 2000) in identifying multisensory integration responses in the superior temporal region that appear to be reflect the binding process and other responses occurring within a distributed network of brain areas that reflect task-related phenomena.

Further support for a role for the STS in the crossmodal binding of auditory and visual linguistic information comes from a recent study using MEG. Raji and colleagues (Raji *et al.*, 2001) examined the neural correlates mediating the integration of auditory (phonemic) and visual (graphemic) aspects of letters (i.e. for stimuli that have been previously associated through

learning). Subjects were scanned whilst exposed to auditory, visual and simultaneously presented audiovisual letters of the Roman alphabet and instructed to identify target letters, regardless of stimulus modality. Audiovisual stimuli included matching and non-matching (i.e. randomly paired) letters. Meaningless auditory (non-speech sounds), visual (false symbols) and audiovisual control stimuli were also presented. Using similar criteria to those employed by Calvert *et al.* (Calvert *et al.*, 2000), group mean audiovisual interaction effects were identified for all stimulus categories in a network of brain areas including the left fronto-parietal region, right frontal cortex, right temporo-occipito-parietal junction (RTOP), and in the right and left STS. Of these, only the RTOP and STS exhibited stronger interaction effects for letters than control stimuli. Stronger interactions to matching than non-matching letters were observed solely in the STS of both hemispheres (greater in the left), suggesting that these sites were 'mainly responsible for the audiovisual integration process'. These data concord with those reported by Calvert *et al.* in highlighting a key role for the STS in the integration of audiovisual linguistic signals (Calvert *et al.*, 2000). Interestingly, the interaction effects in the current study were suppressive [i.e. $AV < (A + V)$], despite faster responses to bimodal stimuli, suggesting that in this experiment, the simultaneous auditory and visual signals inhibited each other or competed within the same neuronal population. The authors speculated that the discrepancy in the direction of interaction effects between the current study and others (Calvert *et al.*, 2000; Callan *et al.*, 2001) may relate to the arbitrariness of the association between auditory and visual letters compared to the intermodal-invariant nature of the frequency-amplitude information contained in audiovisual speech.

Studies Involving Non-linguistic Information. Studies that have attempted to identify putative sites of crossmodal integration using auditory and visual linguistic information are beginning to provide a consistent picture of the brain areas that may be playing a key functional role in this process. One question that arises is whether these areas also play a part in the crossmodal synthesis of stimulus features relating to non-linguistic information. For example, during the integration of simple audible sounds with visual cues in cases where the cross-modal associations have been learned, or for shared physical properties such as temporal synchrony or duration.

Giard and Peronnet used ERPs to investigate the loci and time course of multisensory interactions during the crossmodal integration of feature information for object recognition (Giard and Peronnet, 1999). ERPs were recorded from the scalp whilst subjects identified one of two objects that had been previously defined on the basis of their auditory features (tone A or tone B), visual features (deformation of a circle to an ellipse either in the horizontal or vertical plane) or on the basis of their combined features (object A = tone A + horizontal deformation of the circle and object B = tone B + vertical deformation of the circle). On each trial, subjects were asked to identify object A or B by pressing a key. Behavioural testing both during and prior to scanning found faster reaction times to combined audiovisual stimuli, providing evidence of 'multisensory integration' at the behavioural level. Spatiotemporal analysis of ERPs and scalp current densities revealed several audiovisual interaction components [$AV - (A + V)$] that were temporally, spatially and functionally distinct up to 200 ms post-stimulus onset. These were expressed in both sensory-specific (auditory and visual) and heteromodal brain areas. Interaction effects in visual and auditory cortices resembled predominantly the latency, polarity

and topography of the sensory-specific visual and auditory responses to the unimodal stimuli (the visual P100 and auditory N100 respectively), suggesting multisensory enhancement of, but not necessarily integration in, these areas [but see work by Foxe and colleagues (Foxe *et al.*, 2000)]. In addition, it appeared that these interaction effects were dependent on whether subjects were auditorily or visually dominant. Specifically, interaction effects were greater in the non-dominant sensory-specific cortex, indicating that the inverse effectiveness principle not only applies at the global level of sensory cortex activity in humans but is also dependent on the *subjective* effectiveness of the unimodal cues. Beyond the sensory cortices, the most robust interaction effect was observed in the right fronto-temporal region between 145 and 160 ms after stimulus onset. As this site was not activated by any of the unimodal stimuli, the authors highlighted this region as the locus of multisensory synthesis, at least for the integration of recently learned audiovisual associations. Lack of a sensory dominance effect on interactions in the fronto-temporal cortex led to the further speculation that the different cortical loci may not subserve the same integrative functions. Early interaction effects in sensory-specific cortices were hypothesized to reflect fine and flexible mechanisms that facilitate the processing of the least-efficient cues, fronto-temporal sites carrying out more general integrative functions.

The possibility of multisensory integration in sensory-specific cortices (i.e. early in the cortical processing hierarchy and prior and/or in addition to integration in heteromodal areas) gains further support from a study aimed at identifying the brain areas involved in the synthesis of crossmodal inputs on the basis of their temporal correspondence. Using ERPs, Foxe and colleagues examined the processing stages and source of multisensory interactions between simultaneous but non-associated auditory and somatosensory inputs (Foxe *et al.*, 2000). During recording, subjects were exposed to 1000 Hz tones, stimulated with median nerve electrical pulses, or a combination of the two stimuli presented simultaneously. Spatiotemporal analysis of ERPs and scalp density topographies revealed positive interaction effects at 50 ms after stimulus onset in somatosensory cortex, in the region of the postcentral gyrus. Later interaction effects at 70–80 ms were observed in the region of the posterior auditory cortices in the vicinity of the superior temporal plane. These results are consistent with a model of early integration of the auditory and somatosensory modalities within sensory-specific cortices, at least for the processing of inputs that lack any other association than their shared onsets. Recent data using haemodynamic measures suggest these interactions may be mediated via structures residing deep within the cortical mantle and which are difficult to identify using ERPs.

In a PET study of audiovisual asynchrony detection, in which subjects were explicitly instructed to focus on the temporal coherence of the inputs, Bushara *et al.* (Bushara *et al.*, 2001) identified a rather different pattern of brain areas from that found by Foxe and colleagues. On each trial participants were required to detect whether an auditory tone and a visually presented coloured circle were presented synchronously. Onset asynchrony was varied parametrically, resulting in three levels of difficulty. In a control condition, subjects were asked to decide whether the visual stimulus was green or yellow and to respond only when an auditory stimulus was present. In the control trials, the onsets of the auditory and visual stimuli were always synchronized. Subtraction of the combined asynchronous conditions from the control condition revealed a network of heteromodal brain areas involved in visual-auditory temporal synchrony detection. These included the right insula, posterior

parietal and prefrontal regions. Subsequent regression analysis to identify voxels with regional cerebral blood flow (rCBF) responses that correlated positively with increasing task demand (i.e. decreasing asynchrony) identified a cluster within the right insula, suggesting that this region plays the most prominent role in this process. Interregional covariance analysis also identified task-related functional interactions between the insula and the posterior thalamus and superior colliculus. Together with existing anatomical and electrophysiological data from animal studies, the authors suggested that these findings demonstrate that intersensory temporal processing is mediated via sub-cortical tecto-thalamo-insula pathways.

Similar results have also been reported in a fMRI study of audiovisual temporal correspondence (Calvert *et al.*, 2001). The paradigm was identical to that used to investigate brain areas involved in synthesizing audiovisual speech (Calvert *et al.*, 2000) but in this study, the visual stimulus consisted of an 8 Hz reversing black-and-white checkerboard which alternated every 30 s with a blank screen. The auditory stimulus comprised 1000 ms white noise bursts that were timed either to coincide precisely with the reversal rate of the visual checkerboard (matched experiment) or were randomly shifted out of synchrony (mismatched experiment) whilst maintaining the same number of auditory events. The auditory stimulus alternated with a silent period every 39 s. The structure exhibiting the most significant crossmodal facilitation and suppression to synchronous and asynchronous bimodal inputs respectively was the superior colliculus. Weaker crossmodal interactions were identified in a network of brain areas including the insula/ claustrum bilaterally, the left STS, the right IPS and several frontal regions including the superior and ventromedial frontal gyri. These data concord with those reported by Bushara *et al.* (Bushara *et al.*, 2001) in implicating the superior colliculus and insula in the synthesis of crossmodal cues on the basis of their temporal correspondence. Both studies also suggest that heteromodal cortices play some role in this process, although the nature of their involvement remains to be clarified. The fact that neither study detected the signal changes in sensory-specific cortices reported by Foxe and colleagues (Foxe *et al.*, 2000), despite employing paradigms essentially designed to focus similarly on temporal correspondence, highlights probable differences in the sensitivities of techniques which measure changes in haemodynamics compared to electrical events.

At present, studies of crossmodal identification are beginning to implicate different brain areas in the synthesis of different types of featural information. There is now increasing evidence that the STS plays a key role in the synthesis of audiovisual speech and linguistic signals in general. This is consistent with a growing body of imaging data implicating the human STS in the processing of socially communicative signals [see (Allison *et al.*, 2000)]. Evidence is also emerging that a network incorporating the superior colliculus and insula may be especially sensitive to the temporal correspondence of crossmodal cues. That the consequences of these interactions may be responsible for amplifying the signal intensity in sensory-specific cortices is also becoming increasingly apparent. Where crossmodal associations have been long established (e.g. in the case of speech) these amplifications in early sensory processing areas may be coordinated via back-projections from higher processing areas such as the STS. For more recently learned associations, as in the case of the crossmodal objects reported by Giard and colleagues (Giard *et al.*, 1999), multisensory enhancement of sensory-specific cortices may be mediated initially by areas of frontal cortex as the crossmodal association is explicitly recalled.

Crossmodal Localization

As well as being able to integrate crossmodal inputs relating to stimulus identity, we can also combine spatial coordinate information across senses. This has been referred to as 'crossmodal localization' (Calvert *et al.*, 1998). That different cortical networks may be involved in the crossmodal integration of inputs relating to stimulus identity and stimulus location is suggested by the fact that the McGurk effect (McGurk and MacDonald, 1976) and the ventriloquist's illusion (arising from the crossmodal integration of 'what' and 'where' information respectively) are subject to rather different cognitive constraints (Bertelson *et al.*, 1994). For example, whilst desynchronization of the auditory and visual information has been shown to have deleterious effects on the ventriloquist's illusion, the McGurk effect can still be elicited despite a 180 ms lag between the onset of the auditory and visual inputs (Bertelson *et al.*, 1994).

The neural bases underlying the crossmodal construction of external space have been investigated from two rather different perspectives. One is crossmodal integration, the other, crossmodal spatial attention. The former has involved the use of paradigms in which the crossmodal spatial stimuli are presented sufficiently close in time as to (presumably) induce multisensory integration at the neuronal level. The latter has involved demonstrating that a cue from one modality can speed the detection of a subsequently presented, and spatially concordant, target in another modality. To the extent that such crossmodal attentional enhancements can be elicited even when the cue and target are presented far enough apart in time as to be beyond the time window for multisensory integration at the neuronal level (Meredith *et al.*, 1987) suggests the possibility of two distinct systems involved in crossmodal localization and crossmodal spatial attention [these issues have recently been discussed in greater depth (Macaluso and Driver, 2001; McDonald *et al.*, 2001)]. For the purposes of this review, only studies meeting criteria for crossmodal integration will be described, but their results will be discussed in the light of the current debate concerning the relationship between crossmodal integration and crossmodal spatial attention.

Macaluso and colleagues used fMRI to identify the brain regions involved in co-ordinating the integration of visual and tactile coordinate cues (Macaluso *et al.*, 2000a). Subjects were scanned whilst receiving visual stimuli in the left or right hemifield that were coupled on half of the trials with concurrent tactile stimulation to the right hand. This resulted in spatially concordant visuotactile stimuli when the visual stimulus was presented on the right, and spatially discrepant visuotactile cues when the visual stimulus was presented on the left. By testing for interaction effects between side of visual stimulation and the presence of a tactile input, a significant and spatially specific interaction was identified in the left lingual gyrus. Subsequent effective connectivity analyses were then performed to test for condition-dependent changes in the coupling between brain areas [i.e. the correlation of the response in one area to that measured in a different area (Friston *et al.*, 1997)]. The authors hypothesized that areas mediating the observed crossmodal facilitation in visual cortex should increase their effective connectivity with the left lingual gyrus only during the spatially congruent bimodal stimulation. The most robust effect was observed in the right inferior parietal lobe (TC: $x = 52$, $y = -22$, $z = 34$). This was located in the anterior section of the supra-marginal gyrus. Although multisensory interaction effects were not observed in this area *per se*, the authors suggested that heteromodal areas in the inferior parietal lobe may provide the anatomical substrate for the crossmodal interactions observed in

the lingual gyrus, allowing tactile coordinate information to be transferred to occipital areas via back-projections from an area of heteromodal cortex previously implicated in visuospatial processing.

Lewis and colleagues have investigated brain areas involved in the crossmodal synthesis of dynamic spatial information using fMRI (Lewis *et al.*, 2000). In an initial experiment, subjects performed visual and auditory motion discrimination tasks that involved determining the speed of moving dots or sounds. Each unimodal motion task resulted in a unique pattern of activation extending from the respective primary sensory area to parietal cortex. Superimposition of the two resulting unimodal activation maps revealed areas of co-activation in the lateral parietal cortex (including the IPS), lateral frontal cortex, superior frontal gyrus (SFG) and adjacent AC cortex, and the anterior insula. To discriminate regions of co-responsivity from areas exhibiting multisensory integration, subjects were then scanned during the performance of an explicit crossmodal speed comparison task. Visual and auditory stimuli were presented simultaneously and subjects instructed to decide whether the visual target was moving faster or slower than the auditory target. In two contrasting conditions, subjects were required to ignore the visual stimuli and compare speeds in the auditory domain or vice versa. Comparison of the crossmodal condition with the combined unimodal tasks revealed audiovisual interaction effects predominantly in the IPS (left > right), in the anterior insula and within the anterior midline (SFG and AC). Super-additive responses were also detected in the STS in some individuals, but these responses were typically weak and scattered. These data show that brain areas explicitly involved in the integration of moving auditory and visual stimuli represent a subset of the regions exhibiting co-responsive behaviour. Activation in the intraparietal lobe was considerably more posterior and anterior to that identified by Macaluso and colleagues using static spatial cues (Macaluso *et al.*, 2000a). It is important to note that in the current study, the involvement of heteromodal areas was determined on the basis that these areas exhibited superadditive multisensory integration responses. Attempts to identify brain areas activated solely by the crossmodal task failed to produce any differential activations, suggesting that areas involved in crossmodal integration may also play a role in unimodal processing. This may explain why some electrophysiological and neuroimaging studies that have concentrated on identifying exclusively 'crossmodal' brain areas (Ettinger and Wilson, 1990; Hadjikhani and Roland, 1998) have typically failed to find any such zones.

Heteromodal cortex within the inferior parietal lobe is increasingly being implicated in tasks involving crossmodal localization. The location of these putative integration sites does, however, appear to differ depending on whether the cues are static (Macaluso *et al.*, 2000a) or dynamic (Lewis *et al.*, 2000). To the extent that the IPS has also been implicated in tasks explicitly designed to tap processes involved in coordinating crossmodal spatial attention (Bushara *et al.*, 1999; Eimer, 1999; Macaluso *et al.*, 2000b) it is possible that the crossmodal integration of spatial coordinate information and crossmodal spatial attention may simply be different terms describing the same underlying process.

Crossmodal Learning

Imaging studies investigating the neural mechanisms mediating the development of crossmodal associations have to date been restricted to audiovisual modalities. Using PET, McIntosh and colleagues demonstrated that following a learning period in

which a visual stimulus was consistently paired with an audible tone, presentation of the tone in isolation induced activation in visual cortex (McIntosh *et al.*, 1998). The reverse association was later studied using fMRI (J. Rosenfeld, unpublished). By repeatedly pairing one of two audible tones with a simultaneously presented visual stimulus, subsequent presentation of the visual stimulus alone triggered activation of the primary auditory cortex. These studies demonstrate a role for sensory-specific cortices in the acquisition of novel crossmodal associations and may explain why people report 'hallucinating' the sound or appearance of a stimulus previously paired with a stimulus from another modality even though it is no longer concurrent (Howells, 1994). More recently, Gonzalo and colleagues have attempted to identify time-dependent neural changes related to associative learning across sensory modalities (Gonzalo *et al.*, 2000). Subjects were studied during three separate training sessions, during which time they were exposed to consistently and inconsistently paired audiovisual inputs and to single visual and auditory stimuli. The participants were instructed to learn which audiovisual pairs were consistent over the training period. Time-dependent effects during the acquisition of these crossmodal associations were identified in posterior hippocampus and superior frontal gyrus. Additional activations specific to the learning of consistent pairings included medial parietal cortex and right DLPFC. However, as there were no intramodal (auditory-auditory and visual-visual) associative learning comparison conditions, it is not yet known to what extent the areas identified by these investigators are specific to crossmodal associative processes or reflect generic stimulus-stimulus associative learning operations.

Summary of Findings

The above survey of the imaging literature on crossmodal processing, summarized in Table 1, makes it clear that networks of brain areas, rather than any individual site, seem to be involved in the matching and integration of crossmodal inputs. Components of these networks, however, appear to be differentially specialized for synthesizing different types of crossmodal information (Fig. 2). To date, the role of four such brain areas is becoming increasingly defined. The STS appears to play an important role in the integration of complex featural information, particularly during the perception of audiovisual speech. The IPS, on the other hand, appears to be specialized for synthesizing crossmodal spatial coordinate cues and mediating crossmodal links in attention. Further data suggest that once information from different modalities is blended in these heteromodal sites, the outcome of these interactions is realized by modulation of the signal intensity in sensory-specific cortices via back-projections. The detection of temporal coincidence between crossmodal stimuli appears to be mediated, at least in part, by a predominantly subcortical network including, most prominently, the posterior insula. Whether the temporal signals actually 'converge' in these areas, or whether binding is achieved on the basis of temporal synchronization in the relevant sensory-specific regions, will require further investigation. The involvement of frontal cortex in crossmodal operations is currently less well understood, but there is some evidence that areas within this region may be involved in integrating newly acquired crossmodal associations. Whether the information arriving at these areas is necessarily synthesized in the same way as appears to be the case in the STS and IPS is not clear. Another distinction may relate to the arbitrariness of the crossmodal associations. Whilst the IPS and STS may be involved in the perception and integration of intermodal invariant information,

frontal areas may be recruited when the associations between crossmodal cues are essentially arbitrary. Finally, neuroimaging studies have yet to clarify the relationship between tasks of crossmodal matching and crossmodal integration in terms of the brain areas involved and the mechanisms required to carry out these functions. One area that may have a preferential role in crossmodal matching is the claustrum.

Comments on Methodology

A number of different paradigmatic and analytic strategies have now been used in neuroimaging studies to identify brain areas putatively involved in co-ordinating processing across the senses. To date, the majority of these studies have concentrated on identifying brain areas involved in crossmodal integration. One approach has been to carry out separate experiments using different unimodal stimuli relating to the same object (e.g. the sound of a cat versus the image of a cat) and then seek to identify areas of the brain commonly activated across the experiments using intersections or conjunction analysis (Calvert *et al.*, 1997; Chee *et al.*, 1999; Lewis *et al.*, 2000). To compute an intersection, one simply determines those areas that overlap across the separate unimodal activation maps. Conjunction analysis is a statistical method for determining those brain areas that show 'context-free' activation, i.e. those areas commonly activated across a range of experiments. Friston and colleagues summarized this approach as testing the hypothesis 'that activations across a range of tasks are jointly significant' (Friston *et al.*, 1997). In essence, it consists of combining the data from the separate experiments and determining those areas that show significant activation in the combined data. The statistical inference is therefore performed on the combined data and not on the individual experiments prior to a simple overlapping of maps.

However, there are a number of reasons why computation of conjunctions or intersections between unimodal experiments may not represent an optimal strategy for identifying sites of integration. The first problem relates to the fact that a single voxel in a functional imaging experiment samples a very large number of neurons. Therefore, a significant response to two different unimodal stimuli in the same voxel across separate experiments may simply indicate the coexistence within the voxel of two sets of unimodal neurons responsive to one or other modality, but not both. This strategy may therefore erroneously identify an area as being putatively involved in crossmodal integration (Type 1 error). On the other hand, the requirement for both unimodal stimuli to elicit superthreshold responses if integration areas are to be identified on the basis of overlap, may be overconservative where the response to one of the unimodal stimuli may be weak, as has been shown of some MSI cells in the superior colliculus (Stein and Meredith, 1993). Thus this strategy may also preclude the detection of true multisensory integration sites (Type 2 errors). Any attempt to overcome this problem of weak responses by relaxing the statistical threshold for response detection would immediately run into the problem of Type 1 errors, especially if applied to whole-brain analysis.

One obvious improvement on this strategy would seem to be (at least on the face of it) to contrast unimodal stimuli with true bimodal stimuli, and attempt to identify areas where bimodal stimulation gives a greater response than either modality presented in isolation (Hadjikhani and Roland, 1998; Calvert *et al.*, 1999). For example, in an audiovisual integration experiment, the strategy might be to expose subjects to bimodal stimulation, auditory and visual stimulation and then compute $[AV - V] \cap [AV - A]$. However, if AV is simply the linear sum of A

and V, and the differences are both significant, conjunction analysis may simply detect voxels in which unimodal auditory and visual-responsive neurons coexist. This strategy therefore represents no real improvement on computing the simple intersection $A \cap V$.

A more valid method for identifying integration responses involves the inclusion of a reference (rest) condition in a 2×2 design in which rest, V, A and AV conditions are presented, permitting the computation of the interaction effect $[AV - \text{rest}] - [(A - \text{rest}) + (V - \text{rest})]$. Interaction effects are commonly used in statistical analyses to identify changes that occur when two factors are simultaneously altered that would not be predicted from the results of altering each factor in isolation. In the context of multisensory integration, the use of interaction effects therefore permits the clear demonstration that the bimodal response cannot simply be predicted from the sum of the unimodal responses. This approach therefore addresses directly the problem of voxels containing mixed unimodal populations being erroneously identified as multisensory integration sites, and provides the added advantage of having close analogies with integration behaviour at the neuronal level (see above).

Whilst interaction approaches are thus proving increasingly popular in crossmodal imaging experiments and are demonstrating clear differences from the results of simpler approaches described above [e.g. compare the two papers by Calvert and colleagues (Calvert *et al.*, 1999, 2000)], like any other statistical approach, they must be used and interpreted with caution. For example, superadditive effects to bimodal stimuli may be difficult to detect in the event that the BOLD responses to the unimodal stimuli are at or near ceiling. It may therefore be advisable to study interaction effects within a parametric design in which the intensity of the unimodal stimuli are systematically varied. Serendipitously, the principle of inverse effectiveness may help to counter the problem of BOLD ceiling responses as the largest crossmodal enhancements occur when the unimodal stimuli are minimally effective.

Another issue arises when a stimulus in one modality produces a positive BOLD response in a particular brain area, whereas a different sensory stimulus depresses the BOLD response in this region, below the 'baseline' level. For example, somatosensory tasks have been shown to depress responses in visual cortex (Kawashima *et al.*, 1995), visual tasks to depress responses in auditory cortex (Haxby *et al.*, 1994). In this case, one could inadvertently infer multisensory integration on the basis of positive interaction effects when the two unimodal stimuli are co-presented if the calculation does not take into account (and correct for) the possibility of individual unimodal responses falling below baseline. As mentioned previously, it is not necessary that the individual unimodal responses necessarily reach significance (above baseline) for interaction effects to be identified in response to the bimodal inputs, but it is necessary to ensure that neither of them fall below baseline.

A final issue relates to the interpretation of interaction effects within a fMRI experiment. For example, when two or more brain regions showing interaction effects are detected in the same experiment, one can then ask: are these areas both involved in integration or are some of them simply responding downstream of the integration event itself? It would appear that the only real answer to this question is to obtain accurate timing information on the responses in the different brain regions showing interaction effects (e.g. by using MEG). Thus combined fMRI/MEG experiments yielding good temporal and spatial resolution would appear an excellent approach to resolving this issue. Use of fMRI alone, even utilizing techniques for the

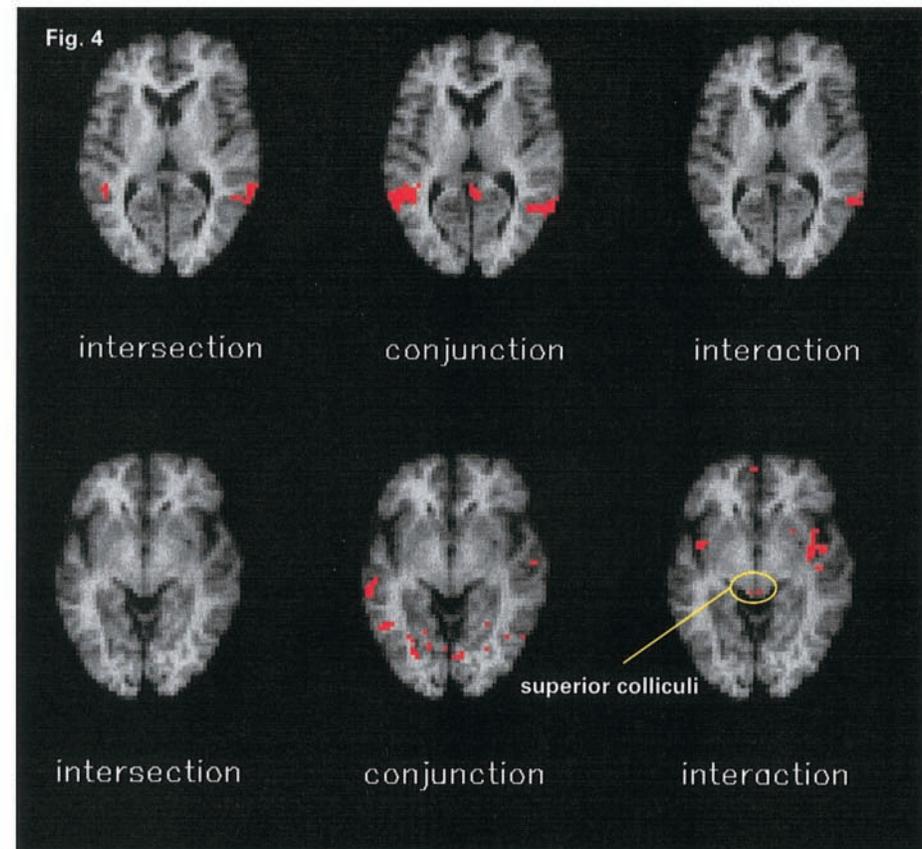
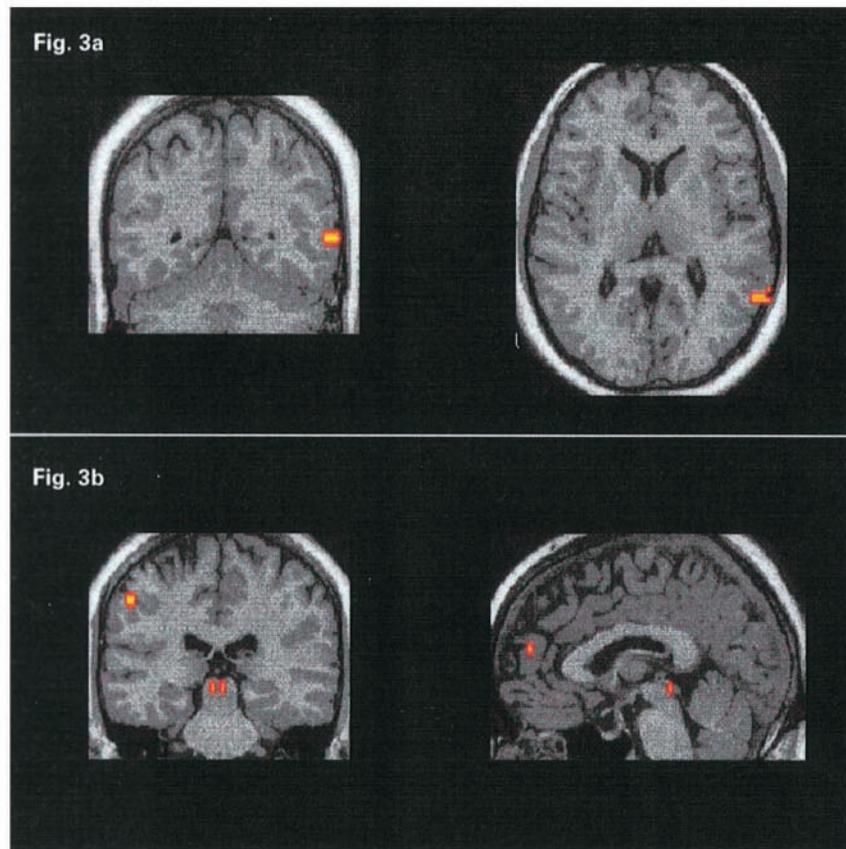


Figure 3. Brain areas showing the strongest positive and negative interaction effects to matched and mismatched speech and non-speech stimuli. (a) Coronal (left, $y = -50$ mm) and axial (right, $z = +9$ mm) sections through the intersection of the positive and negative interaction group effect maps ($n = 10$), both thresholded with a voxel-wise Type 1 error rate of <0.001 and superimposed on a high-resolution structural MRI image warped into the standard stereotaxic space of Talairach and Tournoux (Talairach and Tournoux, 1988). The maximal effect is in the left STS. Activations are shown in radiological convention (i.e. the left side of the image corresponds to the right hemisphere). Reprinted by permission of *Current Biology*. (b) Coronal (left, $y = -30$ mm) and sagittal (right, $x = +1$ mm) sections for the intersection of the non-speech maps, also $n = 10$ and $P \leq 0.001$ at the position of the strongest positive and negative interaction effects. These interactions were identified in the superior colliculi. Reprinted by permission of *NeuroImage*.

Figure 4. The effect of analytic strategy on detection of activated voxels in the axial slices showing the strongest interaction effects. Three analytic methods were employed: (i) intersection of unimodal auditory and visual response maps; (ii) conjunction analysis of the combined bimodal and unimodal data; and (iii) intersection of positive and negative interaction effects. All individual maps were thresholded at a voxel-wise Type 1 error rate of $P < 0.001$ (by randomization testing). The top row shows the maps for the speech stimuli at $z = +9$ mm (Talairach) and the bottom row the non-speech maps at $z = -1$ mm.

estimation of effective connectivity (Macaluso *et al.*, 2000a), appears unlikely to produce such a clear resolution of this problem because such techniques permit estimation of changes in coupling strength but not inference regarding timing issues. There are also major issues in the statistical approaches used to detect significant changes in inter-voxel correlation that underlie studies of effective connectivity (Bullmore *et al.*, 2000).

The Impact of Analytic Strategy on Detection of Multisensory Integration Sites

To illustrate the impact of some of the different analytical strategies discussed above, we can now show the effect of applying them to our own data on the integration of audiovisual speech and non-speech stimuli. The data examined here were obtained using identical experimental paradigms, with only the nature of the stimuli being different (Calvert *et al.*, 2000, 2001). Briefly, the stimuli were auditory speech, visual speech, concordant audiovisual speech, or discordant audiovisual speech (Calvert *et al.*, 2000) or non-speech stimuli in the same combinations (Calvert *et al.*, 2001). In the first experiment (speech stimuli), the main focus of activation (derived from the largest positive and negative interaction effects) was in the left STS (Fig. 3a). In the second experiment (non-speech stimuli), the analogous focus was found in the superior colliculi (Fig. 3b). Having identified these putative integrative sites by the interaction effect methodology, we now examine the ability of the other analytic strategies described above to detect the same brain regions. The results of this comparison are shown in Figure 4. In each experiment, a single axial brain slice is shown which contains the putative integration site showing the strongest effect as determined by the interaction method (Fig. 3a,b). The three analytic strategies that are compared are the following: (i) identify areas showing significant unimodal auditory responses, identify areas showing significant unimodal visual responses and then determine the overlap between the activated areas in the two maps (intersection of superimposed maps); (ii) use the same two sets of data as in (i) but carry out conjunction analysis (Friston *et al.*, 1996); and (iii) identify areas showing significant positive interaction effects with matched auditory and visual inputs, and significant negative interaction effects with mismatched inputs.

Whilst the STS was identified by all three methods in the speech experiment, in the non-speech experiment the superior colliculus is only identified using the interaction effect methodology. It is clear, therefore, that the use of interactions can identify areas of brain distinct from those detected by combination of unimodal data sets (intersection or conjunction) and is not simply a subset of the latter. The main reason for this difference is that analytic methods (i) and (ii) require super-threshold responses to both stimuli, whereas the interaction methodology (iii) does not and will pick up regions that show weak unimodal responses but large (i.e. superadditive) multisensory signal enhancements and response decrements with matched and unmatched stimuli.

The use of interaction effects in experiments on multisensory integration has a number of advantages over the other analytic techniques discussed (but see above for caveats). Firstly, the strategy is based on the known electrophysiological behaviour of cells carrying out signal integration. Secondly, it gives a *de facto* demonstration that integration has occurred as the output signal is significantly different from the sum of the inputs, overcoming the problem that a response to two unimodal inputs could simply be due to different populations of sensory-specific neurons. Thirdly, it allows integrative behaviour to be detected

when unimodal responses are weak. Note that the use of interaction effects necessitates the inclusion of a rest condition as well as both unimodal and bimodal conditions. For studies of crossmodal integration at least, as opposed to crossmodal matching, these conditions will necessarily ensure that the paradigm meets criteria for binding – namely that the two or more temporally proximal sensory inputs are perceived as emanating from a common event (Radeau, 1994).

Conclusions

To date, a number of different imaging techniques have been used to investigate crossmodal interactions in the human brain. The early state of development of the field and the variety of experimental and data analytic approaches employed make it difficult to draw definitive conclusions about the networks of brain areas involved in crossmodal integration and the flow of information through them. Nonetheless, areas of heteromodal cortex, including the STS and IPS, are being consistently implicated during the integration of identity and spatial information respectively. The insula may play a role in the detection of crossmodal coincidence and participate in crossmodal matching. Regions of the frontal cortex may have a more task dependent role in the perception of inputs across multiple modalities.

As the field moves towards more complex questions, e.g. crossmodal integration in the context of emotion processing (Pourtois *et al.*, 2000) and crossmodal mechanisms in pathological states (Surguladze *et al.*, 2001), the main task for researchers in the area at the present time is to establish the most suitable experimental paradigms and analytic methods for the definitive identification of the brain networks involved in crossmodal processing. Combination of data across imaging modalities, and behavioural and electrophysiological techniques, using a well-defined set of paradigms, seems to be the clearest way forward.

Notes

G.A.C. is supported by a Fellowship from the Medical Research Council of Great Britain and the McDonnell-Pew Interdisciplinary Research Centre at the University of Oxford. I would like to thank M.J. Brammer, P.C. Hansen, H. Johansen-Berg and D.M. Lloyd for their helpful comments on this manuscript.

Address correspondence to Gemma A. Calvert, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), John Radcliffe Hospital, Oxford OX3 9DU, UK. Email: gemma@fmrib.ox.ac.uk.

References

- Allison T, Puce A, McCarthy G (2000) Social perception from visual cues: role of the STS region. *Trends Cogn Sci* 4:267–278.
- Banati RB, Goerres GW, Tjoa C, Aggleton JP, Grasby P (2000) The functional anatomy of visual-tactile integration in man: a study using positron emission tomography. *Neuropsychologia* 38:115–124.
- Barth DS, Goldberg N, Brett B, Di S (1995). The spatiotemporal organization of auditory, visual, and auditory-visual evoked potentials in rat cortex. *Brain Res* 678:177–190.
- Baylis GC, Rolls ET, Leonard CM (1987) Functional subdivisions of the temporal lobe neocortex. *J Neurosci* 7:330–342.
- Benevento LA, Fallon J, Davis BJ, Rezak M (1977) Auditory-visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. *Exp Neurol* 57:849–872.
- Bertelson P (1994) The cognitive architecture behind auditory-visual interaction in scene analysis and speech identification. *Cah Psychol Cogn* 13:69–75.
- Bruce C, Desimone R, Gross CG (1981) Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46:369–384.
- Bullmore ET, Horwitz B, Honey G, Brammer M, Williams S, Sharma T

- (2000) How good is good enough for path analysis of fMRI data? *NeuroImage* 11:289–301.
- Bushara KO, Weeks RA, Ishii K, Catalan MJ, Tian B, Rauschecker JP, Hallett M (1999) Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans. *Nat Neurosci* 2:759–766.
- Bushara KO, Grafman J, Hallett M (2001) Neural correlates of auditory-visual stimulus onset asynchrony detection. *J Neurosci* 21:300–304.
- Callan DE, Callan AM, 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 GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS (1997) Activation of auditory cortex during silent lipreading. *Science* 276:593–596.
- Calvert GA, Brammer MJ, Iversen SD (1998) Crossmodal identification. *Trends Cogn Sci* 2:247–253.
- Calvert GA, Brammer M, Bullmore E, Campbell R, Iversen SD, David A (1999) Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport* 10:2619–2623.
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in human heteromodal cortex. *Curr Biol* 10:649–657.
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ (2001) Detection of multisensory integration sites by application of electrophysiological criteria to the BOLD response. *NeuroImage* 14:427–438.
- Chavis DA, Pandya DN (1976) Further observations on corticofrontal connections in the rhesus monkey. *Brain Res* 117:369–386.
- Chee MW, O'Craven KM, Bergida R, Rosen BR, Savoy RL (1999) Auditory and visual word processing studied with fMRI. *Hum Brain Mapp* 7:15–28.
- Desimone R, Gross CG (1979) Visual areas in the temporal cortex of the macaque. *Brain Res* 178:363–380.
- Dodd B (1977) The role of vision in the perception of speech. *Perception* 6:31–40.
- Driver J. (1996) Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* 381(6577): 66–68.
- Duhamel JR, Colby CL, Goldberg ME (1991) In: *Brain and space* (Paillard J, ed.), pp. 223–236. New York: Oxford University Press.
- Eimer M (1999) Can attention be directed to opposite locations in different modalities? An ERP study. *Clin Neurophysiol* 110:1252–1259.
- Ettlinger G, Wilson WA (1990) Cross-modal performance: behavioural processes, phylogenetic considerations and neural mechanisms. *Behav Brain Res* 40:169–192.
- Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cogn Brain Res* 10:77–83.
- Frens MA, Van OA, Van-der WR (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept Psychophys* 57:802–816.
- Fries W (1984) Cortical projections to the superior colliculus in the macaque monkey: a retrograde study using horseradish peroxidase. *J Comp Neurol* 230:55–76.
- Friston KJ, Price CJ, Fletcher P, Moore C, Frackowiak RS, Dolan RJ (1996) The trouble with cognitive subtraction. *NeuroImage* 4:97–104.
- Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ (1997) Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6:218–229.
- Gaffan D, Harrison S (1991) Auditory-visual associations, hemispheric specialization and temporal-frontal interaction in the rhesus monkey. *Brain* 114:2133–2144.
- Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 11:473–490.
- Gonzalo D, Shallice T, Dolan R (2000) Time-dependent changes in learning audiovisual associations: a single-trial fMRI study. *NeuroImage* 11:243–255.
- Gordon B (1973). Receptive fields in deep layers of cat superior colliculus. *J Neurophysiol* 36:157–178.
- Graziano MS, Gross CG (1998) Spatial maps for the control of movement. *Curr Opin Neurobiol* 8:195–201.
- Hadjikhani N, Roland PE (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.
- Harris LR (1980) The superior colliculus and movements of the head and eyes in cats. *J Physiol* 300:367–391.
- Haxby JV, Horwitz B, Ungerleider LG, Maisog JM, Pietrini P, Grady CL (1994) The functional organization of human extrastriate cortex: a PET-rCBF study of selective attention to faces and locations. *J Neurosci* 14:6336–6353.
- Hikosaka K, Iwai E, Saito H, Tanaka K (1988) Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. *J Neurophysiol* 60:1615–1637.
- Howells T (1944) The experimental development of color-tone synesthesia. *J Exp Psychol* 34:87–103.
- Hughes HC, Reuter LP, Nozawa G, Fendrich R (1994) Visual-auditory interactions in sensorimotor processing: saccades versus manual responses. *J Exp Psychol Hum Percept Perform* 20:131–153.
- Ito SI (1982) Prefrontal unit activity of macaque monkeys during auditory and visual reaction time tasks. *Brain Res* 247:39–47.
- Jay MF, Sparks DL (1984) Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature* 309:345–347.
- Jones EG, Powell TP (1970) An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* 93: 793–820.
- Kadunce DC, Vaughan JW, Wallace MT, Benedek G, Stein BE (1997) Mechanisms of within- and cross-modality suppression in the superior colliculus. *J Neurophysiol* 78:2834–2847.
- Kawashima R, O'Sullivan BT, Roland PE (1995) Positron-emission tomography studies of cross-modality inhibition in selective attentional tasks: closing the 'mind's eye'. *Proc Natl Acad Sci USA* 92:5969–5972.
- King AJ, Palmer AR (1985) Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Exp Brain Res* 60:492–500.
- Lewis JW, Beauchamp MS, DeYoe EA (2000) A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb Cortex* 10:888.
- Lewkowicz DJ (2000) The development of intersensory temporal perception: an epigenetic systems/limitations view. *Psychol Bull* 126:281–308.
- Lewkowicz DJ, Lickliter R (1994) The development of intersensory perception: comparative perspectives. Hillsdale, NJ: Erlbaum.
- McDonald JJ, Teder-Salejarvi WA, Ward LM (2001) Multisensory integration and crossmodal attention effects in the human brain. *Science* 292:1791.
- McDonald JJ, Teder-Salejarvi WA, Hillyard SA (2000) Involuntary orienting to sound improves visual perception. *Nature* 407(6806):906–908.
- Macaluso E, Driver J (2001) Spatial attention and crossmodal interaction between vision and touch. *Neuropsychologia* 39:1304–1316.
- Macaluso E, Frith CD, Driver J (2000a) Modulation of human visual cortex by crossmodal spatial attention. *Science* 289:1206–1208.
- Macaluso E, Frith C, Driver J (2000b) Selective spatial attention in vision and touch: unimodal and multimodal mechanisms revealed by PET. *J Neurophysiol* 83:3062–3075.
- McGurk H, MacDonald JW (1976) Hearing lips and seeing voices. *Nature* 264:746–748.
- McIlwain JT (1986) Effects of eye position on saccades evoked electrically from superior colliculus of alert cats. *J Neurophysiol* 55:97–112.
- McIntosh AR, Cabeza RE, Lobaugh NJ (1998) Analysis of neural interactions explains the activation of occipital cortex by an auditory stimulus. *J Neurophysiol* 80:2790–2796.
- Meredith MA, Stein BE (1983) Interactions among converging sensory inputs in the superior colliculus. *Science* 221:389–391.
- Meredith MA, Stein BE (1996) Spatial determinants of multisensory integration in cat superior colliculus neurons. *J Neurophysiol* 75:1843–1857.
- Meredith MA, Nemitz JW, Stein BE (1987) Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors. *J Neurosci* 7:3215–3229.
- Mesulam MM, Mufson EJ (1982) Insula of the old world monkey. III: Efferent cortical output and comments on function. *J Comp Neurol* 212:38–52.
- Mistlin AJ, Perrett DI (1990) Visual and somatosensory processing in the macaque temporal cortex: the role of 'expectation'. *Exp Brain Res* 82:437–450.
- Mufson EJ, Mesulam MM (1984) Thalamic connections of the insula in the

- rhesus monkey and comments on the paralimbic connectivity of the medial pulvinar nucleus. *J Comp Neurol* 227:109-120.
- Murray EA, Mishkin M (1985) Amygdalectomy impairs crossmodal association in monkeys. *Science* 228:604-606.
- Murray EA, Malkova L, Goulet S (1998) In: *Comparative neuropsychology* (Milner AD, ed.), pp. 51-69. Oxford: Oxford University Press.
- O'Hare JJ (1991) Perceptual integration. *J Wash Acad Sci* 81:44-59.
- Pandya DN, Yeterian EH (1985) In: *Cerebral cortex* (Peters, A, Jones EG, eds), pp. 3-61. New York: Plenum Press.
- Pearson RC, Brodal P, Gatter KC, Powell TP (1982) The organization of the connections between the cortex and the claustrum in the monkey. *Brain Res* 234:435-441.
- Peck CK (1987) Visual-auditory interactions in cat superior colliculus: their role in the control of gaze. *Brain Res* 420:162-166.
- Perrott DR, Saberi K, Brown K, Strybel TZ (1990) Auditory psychomotor coordination and visual search performance. *Percept Psychophys* 48:214-226.
- Pourtois G, de Gelder B, Vroomen J, Rossion B, Crommelinck M (2000) The time-course of intermodal binding between seeing and hearing affective information. *NeuroReport* 11:1329-1333.
- Radeau M (1994) Auditory-visual spatial interaction and modularity. *Curr Psychol Cogn* 13:3-51.
- Raij T, Uutela K, Hari R (2001) Audiovisual integration of letters in the human brain. *Neuron* 28:617-625.
- Reisberg D (1987) In: *Hearing by eye: the psychology of lip-reading* (Dodd B, Campbell R, eds). Hillsdale, NJ: Lawrence Erlbaum.
- Rizzolatti G, Scandolara C, Matelli M, Gentilucci M (1981a) Afferent properties of periarculate neurons in macaque monkeys. I. Somatosensory responses. *Behav Brain Res* 2:125-146.
- Rizzolatti G, Scandolara C, Matelli M, Gentilucci M (1981b) Afferent properties of periarculate neurons in macaque monkeys. II. Visual responses. *Behav Brain Res* 2:147-163.
- Rose SA, Ruff HA (1987) In: *Handbook of infant development*, 2nd edn (Osofsky J, ed.), pp. 318-362. New York: Wiley.
- Sams M, Aulanko R, Hamalainen M, Hari R, Lounasmaa OV, Lu ST, Simola J (1991) Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neurosci Lett* 127:141-145.
- Seltzer B, Pandya DN (1978) Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res* 149:1-24.
- Seltzer B, Pandya DN (1980) Converging visual and somatic sensory cortical input to the intraparietal sulcus of the rhesus monkey. *Brain Res* 192:339-351.
- Singer W, Gray C (1995) Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci* 18:555-586.
- Stein BE, Meredith MA (1993) *Merging of the senses*. Cambridge, MA: MIT Press.
- Stein BE, Magalhaes-Castro B, Kruger L (1975) Superior colliculus: visuotopic-somatotopic overlap. *Science* 189(4198):224-226.
- Stein BE, Huneycutt WS, Meredith MA (1988) Neurons and behavior: the same rules of multisensory integration apply. *Brain Res* 448:355-358.
- Stein BE, Meredith MA, Huneycutt WS, McDade L (1989) Behavioural indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *J Cogn Neurosci* 1:12-24.
- Sumbly WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26:212-215.
- Surguladze SA, Calvert GA, Brammer MJ, Campbell R, Bullmore ET, Giampietro V, David AS (2001) Audio-visual speech perception in schizophrenia: an fMRI study. *Psychiat Res* 106:1-14.
- Talairach J, Tournoux P (1988) *Co-planar stereotactic atlas of the human brain*. Stuttgart: Thieme.
- Turner BH, Mishkin M, Knapp M (1980) Organization of the amygdalopetal projections from modality-specific cortical association areas in the monkey. *J Comp Neurol* 191:515-543.
- Vaadia E, Benson DA, Hienz RD, Goldstein MHJ (1986) Unit study of monkey frontal cortex: active localization of auditory and of visual stimuli. *J Neurophysiol* 56:934-952.
- Wallace MT, Meredith MA, Stein BE (1992) Integration of multiple sensory modalities in cat cortex. *Exp Brain Res* 91:484-488.
- Wallace MT, Meredith MA, Stein BE (1993) Converging influences from visual, auditory, and somatosensory cortices onto output neurons of the superior colliculus. *J Neurophysiol* 69:1797-1809.
- Wallace MT, Wilkinson LK, Stein BE (1996) Representation and integration of multiple sensory inputs in primate superior colliculus. *J Neurophysiol* 76:1246-1266.
- Wilkinson LK, Meredith MA, Stein BE (1996) The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Exp Brain Res* 112:1-10.
- Zahn JR, Abel LA, Dell'Osso LF (1978) Audio-ocular response characteristics. *Sens Processes* 2:32-37.