Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Physiology - Paris 98 (2004) 191–205
Journal of Physiology Paris
www.elsevier.com/locate/jphysparis

Multisensory integration: methodological approaches and emerging principles in the human brain

Gemma A. Calvert *, Thomas Thesen

University Laboratory of Physiology, University of Oxford, Parks Road, Oxford OX1 3PT, UK

Abstract

Understanding the conditions under which the brain integrates the different sensory streams and the mechanisms supporting this phenomenon is now a question at the forefront of neuroscience. In this paper, we discuss the opportunities for investigating these multisensory processes using modern imaging techniques, the nature of the information obtainable from each method and their benefits and limitations. Despite considerable variability in terms of paradigm design and analysis, some consistent findings are beginning to emerge. The detection of brain activity in human neuroimaging studies that resembles multisensory integration responses at the cellular level in other species, suggests similar crossmodal binding mechanisms may be operational in the human brain. These mechanisms appear to be distributed across distinct neuronal networks that vary depending on the nature of the shared information between different sensory cues. For example, differing extents of correspondence in time, space or content seem to reliably bias the involvement of different integrative networks which code for these cues. A combination of data obtained from haemodynamic and electromagnetic methods, which offer high spatial or temporal resolution respectively, are providing converging evidence of multisensory interactions at both "early" and "late" stages of processing—suggesting a cascade of synergistic processes operating in parallel at different levels of the cortex.

Keywords: Multisensory integration; FMRI; Imaging; MEG

1. Introduction

The past decade has witnessed a growing shift of emphasis away from the study of the senses in isolation and towards an understanding of how the human brain coordinates the unique sensory impressions provided by the different sensory streams. The adoption of a multisensory perspective on human sensory perception has evolved in part as a consequence of developments in both technology and sensory neurophysiology. In the late 1980s and 1990s, the introduction of novel brain imaging techniques such as positron emission tomography (PET), functional magnetic resonance imaging (FMRI) and magnetoencephalography (MEG) allowed, for the first time, the study of global brain function in vivo. One consequence of this development was that research could now focus on how systems interacted, rather than how they behaved in isolation. These advances in technology coincided with a time of increasing knowledge about the mechanisms involved in the primary sensory systems. A natural extension of this understanding was the realization that a complete understanding of our perceptual systems would necessitate the inclusion of how each sense was modulated by or integrated with input arriving from different sensory systems.

The evolutionary basis of such multisensory capabilities is clear. Integrating inputs from multiple sensory sources disambiguates the discrimination of external stimuli and can speed responsiveness (see [91] for a review). The question that now confronts us is how best to study these phenomena in the human brain. What are the opportunities afforded by the different techniques and what kinds of strategies should we employ to tease out the key principles, some of which may be unique to humans? Specific questions that are currently being addressed using human neuroimaging methods (often in conjunction with single cell recording studies in non-human primates) include (i) what is the nature of the neuronal mechanisms mediating multisensory

---

[*] Corresponding author.
E-mail address: gac@physiol.ox.ac.uk (G.A. Calvert).

0928-4257/$ - see front matter © 2004 Published by Elsevier Ltd.
doi:10.1016/j.jphysparis.2004.03.018

integration (ii) where are these neuronal networks localized (iii) are distinct networks involved in synthesizing different types of information such as time, spatial location, content (iv) at what stage of processing are these integrative computations being carried out (i.e. "early" versus "late" integration) and (v) how best can each of these questions be examined in the human brain?

In this paper, we first provide a very brief overview on what is currently known about multisensory integration based on behavioral studies in humans and neuroanatomical and electrophysiological studies in monkeys. These areas have been reviewed extensively elsewhere (see [12,22,45,90,105]). We will then provide a concise description of currently available neuroimaging techniques and their relative merits. This is followed by a discussion of the various imaging paradigms and analytic strategies that have so far been utilized in the investigation of multisensory phenomena, and the advantages and disadvantages of different approaches. These studies are now beginning to implicate certain brain areas in the crossmodal synthesis of different stimulus parameters such as time, space and identity—and are briefly summarized here (for more detailed reviews of current findings, see [16,18,57,78]). We then highlight a topical issue in the multisensory literature—that of the role of endogenous and exogenous attentional processes in the context of crossmodal binding. Finally, we consider what is special about multisensory convergence and conclude with some suggestions concerning future research directions.

1.1. Behavioral studies

In an early study of crossmodal phenomena, [96] demonstrated that reaction time (RT) in a target detection task can be speeded by the presence of a non-specific accessory stimulus in another modality, i.e. a stimulus that bears no meaningful relationship other than temporal proximity. Subsequent investigations into this crossmodal 'redundant target effect' (RTE) have replicated and extended these findings [4,21,33,43,85] and provided further evidence that the observed crossmodal facilitation is not simply due to a statistical probability summation effect alone [34,69]. Consequently, "race models" of the RTE that sought to explain the phenomenon on the basis of a probabilistic interpretation, have been largely superceded by "co-activation models" [66] in which signals from the different sensory channels are integrated prior to initiation of the motor response.

Behavioral studies have also explored the conditions under which crossmodal interactions occur. Two key determinants of intersensory binding are synchronicity and spatial correspondence [76]. Thus, when two or more sensory stimuli occur at the same time and place, they are typically bound into a single percept and detected more rapidly than either input alone. In contrast, slight discrepancies in the onset and location of two crossmodal cues can be significantly less effective in eliciting responses than isolated unimodal stimuli [86,93]. Similar instances of crossmodal facilitation have also been shown to effect detection thresholds. For example, Frassinetti et al. [30] found that subject's sensitivity to visual stimuli presented below luminance threshold was increased by a simultaneous accessory sound burst presented at the same spatial location. This effect was eliminated when the two sensory inputs were separated in space or offset by more than 500 ms. Similar crossmodal influences have also been reported in the case of auditory and tactile detection thresholds (for reviews see [55,105]).

In addition to the parameters of time and space, psychophysical experiments have shown that the synthesis of multisensory inputs can also be influenced by their semantic congruence. For example, hearing a dog's bark emanating from the same approximate location as a visible cat is unlikely to create the impression of a barking cat. On the other hand, multisensory inputs concerning object identity can be combined to produce a novel perceptual outcome, one that was neither heard nor seen. Dubbing an audible syllable (BA) onto videotape of a speaker mouthing a different syllable (GA) typically results in the perception of "DA" [62]. Because the contextual information from the auditory and visual channels is complementary and persuasive, the effect can tolerate temporal and spatial disparity to a greater degree than two simple inputs that have no shared content-related information.

In addition to differences in the physical properties of the stimuli, it is beginning to come to light that other factors too may play a role in mediating crossmodal interactions. These include task-related factors, such as attended modality and whether subjects are required to detect or discriminate a target [53], as well as other intrinsic variables such as the prior sensory bias of the subject (i.e. whether they are visually or acoustically dominant [32]).

1.2. Neuroanatomical findings

In the late 1960s and 1970s, it was widely accepted that cortical sensory processing progressed in a hierarchical fashion from primary to secondary sensory-specific cortices to regions of "association" or "heteromodal" cortex. These so-called "heteromodal" zones were defined on the basis that they were found to receive converging afferents from multiple sensory modalities and contained neurons responsive to stimulation in more than one modality. Studies carried out during that period, and more recently, have identified a large number of such areas (see Fig. 1). These include anterior portions of the superior temporal sulcus (STS)

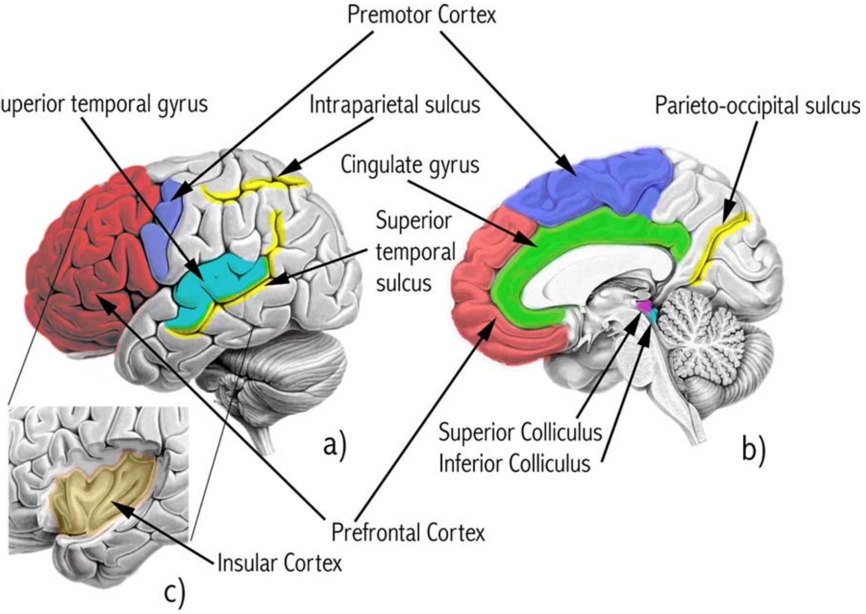

Fig. 1. Lateral (a) and mid-sagittal (b) views of the human brain showing putative heteromodal brain areas. (c) Shows insular cortex after temporal lobe dissection. Different regions of heteromodal cortex are depicted in distinct colours across lateral and medial views. Yellow defines the boundaries of multisensory regions implicated in cortical sulci. Delineation of these areas has been based on neuroanatomical, electophysiological and laminar profile studies in non-human primates (see Section 1.2 for a detailed explication of these brain regions).

[2,3,8,19,71,103] posterior portions of the STS, including the temporo-parietal association cortex (Tpt) [20,49] parietal cortex, including the ventral (VIP) and lateral (LIP) intraparietal areas [7,51,52], and premotor and prefrontal cortex ([36,104]). Multisensory convergence zones have also been identified in sub-cortical structures, including the superior colliculus [31], the claustrum [74], the suprageniculate and medial pulvinar nuclei of the thalamus [64,70], and within the amygdaloid complex [98].

Although a strictly hierarchical view of sensory processing has been challenged by more recent evidence indicating a more divergent and parallel organization (for a review see [65]), the relative synaptic distance of these heteromodal zones from regions of primary sensory cortex still bolsters the prevailing view that multisensory integration occurs at a "late" stage of processing, following considerable elaboration of the unisensory signals in their respective "dedicated" cortices. However, evidence from a number of sources suggests that such a model may be over-simplistic. For example, systematic lesions of large swathes of so-called heteromodal cortex in monkeys have failed to produce reliable deficits on tasks of crossmodal matching and transfer (reviewed in [25]) and electrophysiological studies in humans have found evidence of interaction effects as early as 40 ms post-stimulus onset, consistent with interaction of the two sensory modalities at a very early stage of processing (e.g. [32]).

The debate between "late" and "early" models of multisensory integration has taken on greater significance following recent evidence that areas early in the cortical auditory processing hierarchy project directly to areas early in the visual hierarchy, including V1 [25,81]. For example, using retrograde tracers, Falchier et al. [26] identified projections from core and parabelt auditory cortex into parts of V1 corresponding to the representation of the peripheral visual field. Whilst the distribution of these connections was relatively sparse, they nevertheless provide a possible mechanism by which the auditory system could alert visual cortex to an expected visual stimulus. Evidence of early multisensory interactions within putative "unisensory" cortex has also been reported recently. In a series of elegant studies examining the time course and laminar profile of somatosensory, visual and auditory inputs into posterior auditory cortex (belt and parabelt regions) Schroeder et al. [83,84] observed multisensory integration responses in these auditory-responsive areas at very short latencies. Recording of laminar response profiles identified that in auditory cortex, somatosensory and auditory inputs have a feedforward pattern whilst visual inputs to the region have a feedback pattern. Although the source of these somatosensory and visual inputs into auditory cortex remains to be determined, evidence from tracer studies (as outlined above) suggest a pattern of projections from both heteromodal and unimodal cortices.

To summarize, the most current indications from neuroanatomical studies in monkey suggest that multisensory integration could be achieved at various levels of the cortical processing hierarchy. Thus, integration of the different sensory streams could occur at both early and late stages of processing mediated via a parallel network of both feedforward and feedback connections. We have now outlined several routes by which the senses might converge. Electrophysiological studies are beginning to detail the neuronal mechanisms that might implement these putative synergistic processes.

1.3. Multisensory integration at the cellular level

The most detailed studies of crossmodal interactions at the neuronal level have been conducted in the mammalian superior colliculus (SC) (see [91]). Single-unit recordings from this subcortical structure, which is thought to be involved in orientation and attentive behaviors, suggest certain neuronal mechanisms and rules by which multisensory convergence is achieved. For example, multisensory neurons in the SC display overlapping sensory receptive fields, one for each modality (A, V, T) to which they respond. When two or more sensory cues occur in close temporal and spatial proximity, the response of these neurons can be substantially enhanced, sometimes exceeding 12-fold enhancements in firing rate beyond that expected by summing the impulses exhibited by each unimodal input in isolation [91]. Because the output no longer resembles the response obtained to either input, there is a de facto assumption that the information obtained from two sources has been combined to form a single (new) output signal [94]. This process is referred to as *multisensory integration*. The observed facilitation of the neuronal response is often maximal when the responses to the individual inputs are weakest, a principle known as *inverse effectiveness*. In contrast, crossmodal stimuli that show spatial or temporal disparity can induce profound *response depression*. This means that the response to an unimodal stimulus can be severely lessened, even eliminated, by the presence of an incongruent stimulus from another modality [46]. These principles of multisensory integration have also been shown to apply to superior colliculus-mediated functions such as orientation and attentive behaviors [89,92] as well as a range of other crossmodal interactions that may be subserved by other brain areas.

Apart from the SC, neurons exhibiting multisensory receptive fields have also been shown to be present in cortical structures of the monkey [23,24,35,68], cat [106] and rat [1]. However, detailed observations of multisensory response properties in cortex are comparably sparse and sometimes vary from those of the SC. For example, in the cat multisensory integration responses in neurons of the anterior ectosylvian fissure and the lateral sulcus were less restrained by the precise temporal and spatial congruency of the multisensory stimuli [102]. This suggests that multisensory processing in the cerebral cortex may subserve different and a wider range of functions, most of which remain to be explored.

Human neuroimaging techniques now offer us an avenue by which to explore both the routes and mechanisms of multisensory integration in humans. In this next section, we will evaluate briefly some of the techniques available for investigating these crossmodal phenomena. In the subsequent section, we go on to discuss some of the methodological and analytic strategies that have been adopted using these techniques to investigate multisensory processes, the pros and cons of different approaches and the assumptions that they incorporate.

2. Neuroimaging methods

The various neuroimaging methods that have been used in the investigation of human multisensory brain mechanisms fall into two categories:

1. haemodynamic/metabolic, of which the most prominent techniques are functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) and
2. electrical/magnetic, which includes electroencephalography (EEG) and magnetoencephalography (MEG).

These imaging techniques differ, not only in terms of their temporal and spatial resolution (Fig. 2), but also in the source of their respective signals which can have profound consequences on the interpretation of the data obtained using these methods. In the following section we provide a brief description of these techniques and the basis of their signals.

2.1. Haemodynamic methods

Haemodynamic neuroimaging methods rely on the assumption that task-induced neuronal activity is related to changes in both local cerebral blood flow and oxygen metabolism. These changes in the circulatory system in the region of neuronal activation can be used to derive inferences about the underlying neuronal activity and are therefore *indirect* measures of that activity. Of these methods, PET and BOLD FMRI have been most commonly applied to imaging multisensory processes in the human brain.

2.1.1. PET

PET allows the measurement of changes in neural activity by monitoring task-related changes in regional cerebral blood flow (RCBF) or corresponding blood volume (RCBV) [42,77]. During PET scanning, a radioactive solution containing positron-emitting atoms is introduced to the blood stream. These positrons interact with electrons to produce photons of electro-

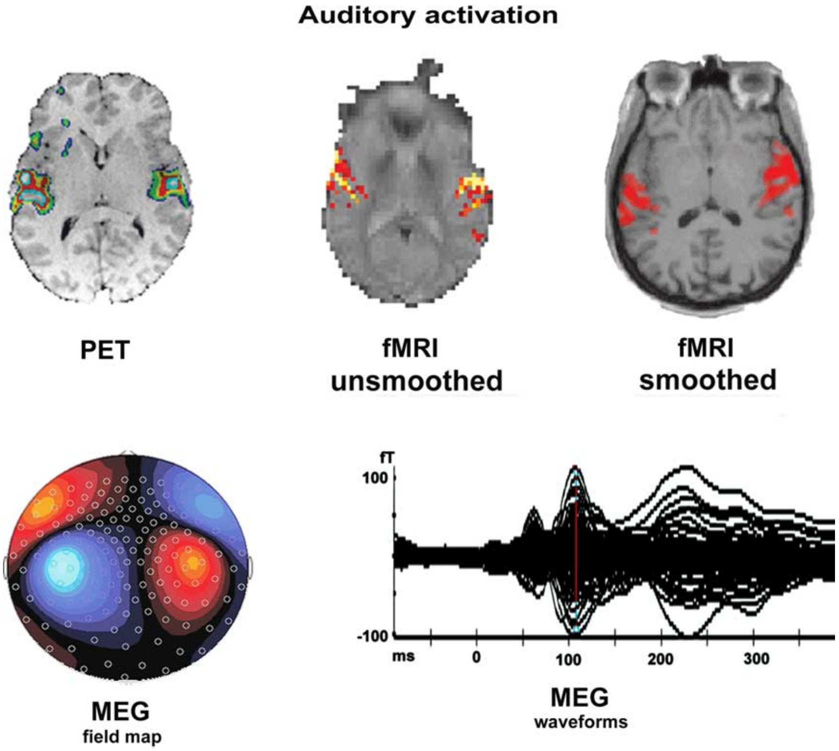

Fig. 2. Provides a comparison of the relative spatial resolution and data format offered by PET, FMRI and MEG. Activations in each case were obtained from a single individual in response to simple auditory stimuli. The PET data was acquired at the University of Melbourne as part of a study on auditory hallucinators (collaborative data supplied by Dr. M. Seal, Department of Psychiatry, Institute of Psychiatry, UK). Red and indigo reflect the peaks of activation and the data were superimposed onto a high-resolution T1 MRI image of the individual's brain. The superior spatial resolution of FMRI can be observed both in the unsmoothed and smoothed images. The unsmoothed FMRI data were acquired on a 1.5T MRI scanner at the Institute of Psychiatry, London, UK by Professor M. Brammer. Yellow reflects voxels of highest statistical significance. The data are superimposed onto the individual's brain acquired at the same resolution as the FMRI data. The activation clearly delineates the transverse temporal gyri (one in the left hemisphere and two gyri in the right hemisphere) or primary and parabelt auditory regions with considerably better spatial resolution than the PET image. The second FMRI (smoothed data superimposed onto a higher resolution T1 image of the subject's brain) image shows bilateral activation encompassing HG and the surrounding STP. Activation is more anterior on the right consistent with the asymmetrical anatomy of Heschls gyrus Data are from subject 4 reported by Hall et al. [107]. The activation map is for the statistical contrast between pairs of matched frequency-modulated and unmodulated tones ($P < 0.001$ uncorrected). The MEG images are generated from the same data set and show the responses to a simple auditory stimulus recorded with a 151 channel MEG system (courtesy of the Wellcome Centre for MEG Studies, University of Aston, UK). The rightmost MEG image shows the timecourse of the individual responses overlaid to each other to generate the evoked potential waveform. This gives an example of the comparatively superior temporal resolution that MEG offers over haemodynamic methods. The leftmost MEG image displays a field map obtained 100 ms after stimulus presentation. Yellow–red contours show the magnetic field distribution on the skull surface and illustrates the relatively poorer spatial resolution of MEG. The field map activity pattern corresponds to the N100 m component of the auditory evoked potential, which is assumed to originate from the supratemporal auditory cortex [40].

magnetic radiation, whose trajectory can be determined and reconstructed by a ring of photon detectors around the subject's head. By comparing images of blood flow under different experimental conditions, brain regions implicated in each component of the task can be isolated. The spatial resolution of PET is in the order of 4–6 mm and its temporal resolution, which is constrained by the image acquisition time, is between 90 and 120 s.

2.1.2. FMRI

Over the last decade, FMRI has arguably overtaken PET as the method of choice for studying brain function in vivo. This is largely due to the fact that FMRI does not require injection of a radioactive substance, but relies instead on a natural contrast agent—the blood oxygen dependent level (BOLD) effect [47,72]. BOLD contrast exploits the different magnetic properties of oxygenated and deoxygenated blood. When a subject is placed in a high magnetic field (the MR scanner), task-induced changes in brain metabolism (see above) alter the ratio of oxy- and deoxy-hemoglobin locally causing measurable changes in MR signal intensity [47,72,99]. Detection of these MR signals changes allows the source of the underlying neuronal activity to be localized to within a few millimeters. While this provides FMRI with very high spatial resolution, its temporal resolution is restricted by the sluggish nature of the BOLD response (approximately 6–8 s post-stimulus onset). Recent data has shown that the main source of the BOLD signal derives from changes in local field potentials, rather than neural spike activity [54]. In other words, BOLD-FMRI detects areas of the brain that process information arising from a stimulus, rather than neurons that fire to respond directly to that stimulus.

2.2. Electrical recordings (EEG/MEG)

In contrast to PET and FMRI, EEG and MEG provide a direct measure of the electrical or electromagnetic fields generated by neuronal activity. MEG and EEG signals, typically recorded from the surface of the skull, are thought to reflect the synchronous activation of large neuronal populations; more specifically, from post-synaptic trans-membrane currents in pyramidal dendrites. The transmission of the neuronal current through the brain and to the sensors is virtually instantaneous and thereby provides EEG and MEG with a temporal resolution on a millisecond timescale. Although the temporal resolution of these electromagnetic methods are clearly superior to haemodynamic techniques, their spatial resolution is somewhat restricted by a technical limitation which requires the calculation of a signal source in a three-dimensional object (i.e. the subject's head) based on the two-dimensional information provided by the sensors.

In MEG and EEG studies the activity related to a specific sensory event of interest is extracted from the background activity by averaging over multiple event trials, time-locked to the stimulus onset. The resulting waveform pattern, the evoked response, is comprised of a deflection of positive and negative peaks over time. These evoked responses are then typically modeled as current dipoles whose number, strength and locations are estimated based on the externally measured electrical or magnetic field distribution. This procedure poses a non-trivial challenge because there is no mathematically unique solution to the problem of inferring the numbers and locations of dipoles that could, theoretically, produce the observed pattern of activity on the surface of the skull. In other words there is an infinite number of source configurations that could produce exactly the same measured field—the so-called *inverse problem* [41]. Solving the inverse problem for one local source of activation is straightforward since the laws governing the propagation of electrical/magnetic fields through skull and tissue, the so-called *forward problem*, are now well understood [73]. However, if several sources are contributing to the observed field map, as can be expected in multisensory experiments, solving the inverse problem becomes increasingly more difficult and complex. In practice, however, the experimenter can use a priori knowledge of physiology and functional anatomy, often derived from other neuroimaging modalities, to incorporate feasible constraints into the model [17,39].

3. Neuroimaging paradigms and analytic strategies

Several different strategies have been used to identify brain areas involved in crossmodal interactions in humans. At present, these can be most usefully divided by task and paradigmatic/analytic approach. For example, some researchers have employed tasks of *crossmodal matching* that involve the explicit comparison of information from different modalities and pertaining to two distinct objects (e.g. [37,38]). Another simple approach has been to compare brain activity evoked during presentation of information in one modality (e.g. auditory) to that evoked by the same (or related) task carried out in another modality (e.g. visual), where a *superimposition* of the two maps reveals areas of common (or 'co-responsive') activation ([10,59]). Others have utilized paradigms explicitly designed to tap *crossmodal integration* ([14,15,29]). Here, information from two different modalities perceived as emanating from a common event is fused into an integrated percept [76].

The results of these studies are beginning to challenge the view posited by Ettlinger [25] that all crossmodal phenomena "require only one underlying process". Rather, these experiments have shown that different networks of brain areas are involved in different crossmodal tasks and may be further differentiated based on the variable being manipulated (spatial, temporal or featural correspondence) and the particular combination of senses under study (for a review, see [12]).

We will now discuss the advantages and disadvantages of the different experimental and analytic strategies used to identify possible sites of multisensory convergence. In addition, we highlight some of the assumptions associated with each approach, and the additional caveats that have to be taken into account depending on the particular imaging methodology applied. As will become clear, different paradigms and different imaging techniques can be used to address distinct questions concerning multisensory processing such as the loci of multisensory convergence, the question of whether integration occurs at early or late stages of processing, and whether these mechanisms are modulated by factors such as task and attention. We will illustrate our points in the context of auditory (A) and visual (V) and tactile (T) inputs.

3.1. Haemodynamic studies

The superior temporal resolution of PET and FMRI coupled with their relatively poor timing information, makes these techniques best suited to the localization of putative multisensory convergence sites. Thus, although both methods can be used to identify interactions and their source, they provide sparse information concerning the timecourse and route of multisensory interactions.

Several different strategies for localizing these integration sites have been adopted.

3.1.1. Superimposition of two unimodal tasks

One obvious approach to identifying multisensory brain areas has been to expose subjects to stimuli presented in one or another modality and localize the brain areas responsive to both. In the context of an FMRI or PET experiment, this typically involves superimposition of the two unisensory activation maps in standard stereotactic space and subsequently determining area(s) of overlap. There are several advantages to such an approach. Firstly, experiments of this nature are relatively straightforward to design providing the task is equally relevant in the two different sensory domains (but see below). Secondly, they have a certain face validity in that most multisensory neurons recorded in animals are defined on the basis that they are responsive to stimuli in more than one modality (see [65,91]). Indeed, using this strategy, several human imaging studies (e.g. [50,59,78]) have localized sites of multisensory co-responsiveness in areas believed to be homologous to those previously identified as bimodal using electrophysiological techniques in non-human primates.

Although superimposition is one useful method for identifying areas responsive to tasks or stimulation experienced in more than one modality, it does not allow us to draw conclusions concerning the presence of bimodal (i.e. those that respond to both modalities) cells in co-activated areas per se. Neither is it possible to demonstrate that these areas necessarily *integrate* multisensory cues. To explain, superimposing the activation maps from an auditory and a visual object recognition task and determining the regions of overlap will in theory (but see below) isolate brain areas responsive to the auditory and visual stimuli as well as those involved in processing the non-sensory components of the paradigm (e.g. working memory). Discriminating these components would thus require further experiments. Even under these circumstances, co-responsiveness (as determined by overlap) may simply indicate either (a) the presence of mixed populations of unisensory cells within activated voxels, but none that are actually bimodal, or (b) the presence of bimodal cells that are co-responsive but fail to integrate their inputs. For example, electrophysiological studies in the superior colliculus have found that although many neurons in this structure are co-responsive to multiple modalities, not all of them necessarily combine or integrate this information such that their response to one modality is measurably altered by the presence of a stimulus from a different modality (“*multisensory integration*”—*see above*).

3.1.2. Bimodal versus unimodal contrasts

A possible improvement on the superimposition methodology is to contrast activation obtained during unimodal tasks against that produced during actual bimodal stimulation, and attempt to identify areas where bimodal stimulation gives a greater response than either modality presented in isolation (e.g. [13,37]). Here the idea is to expose subjects to bimodal stimulation, auditory and visual stimulation and then compute the conjunction $[AV - V] \cap [AV - A]$. Ostensibly, this strategy benefits from the obvious inclusion of a bimodal condition that may invoke synergistic processes not similarly recruited during unimodal stimulation. Furthermore, the detection of conjunction responses allows the extraction of context-independent bimodal activity (i.e. irrespective of whether the bimodal–unimodal contrast was calculated against a purely visual or purely auditory condition). However, if AV is simply the linear sum of A and V, and the differences are both significant, conjunction analysis may still simply detect voxels in which unimodal auditory and unimodal visual-responsive neurons coexist. Thus, this strategy may not afford any real improvement over computing the simple intersection $[A \cap V]$.

A more robust method for identifying integration responses involves the inclusion of a reference (“rest”) condition in a full 2×2 design in which rest, A, V, and AV conditions are all present, permitting the computation of the interaction effect $[AV\text{-rest}] - [(A\text{-rest}) + (V\text{-rest})]$. Interaction effects are commonly used in statistical analysis 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 has a number of advantages over the other analytic techniques discussed. First, the strategy is based on the known electrophysiological behavior of cells carrying out signal integration. Second, it provides a de facto demonstration that some form of interaction 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. Third, it allows integrative behavior to be detected when unimodal responses are weak. For studies of crossmodal integration at least, these conditions will necessarily ensure that the paradigm meets criteria for binding, such that the two or more temporally proximal sensory inputs are perceived as emanating from a common event [76]. Finally, because the calculation of interaction effects requires the inclusion of both unisensory, a multisensory and rest condition, it is also possible to compute superimpositions and conjunctions for comparison.

However although we believe interaction effects currently permit the strongest conclusions concerning multisensory convergence to be drawn, they are not immune from issues of interpretation. For example, Laurienti et al. [48] have demonstrated that putative multisensory integration responses ($AV > A + V$) could arise as a consequence of summing positive and negative BOLD responses to stimuli in a single modality. In a recent FMRI study, they observed that when subjects received visual stimulation, activation in the auditory cortex was suppressed below the resting BOLD baseline (i.e. de-activated). The same effect was also observed in the opposite direction (i.e. in the visual cortex) during acoustic stimulation. Thus, when calculating whether the bimodal stimulation was greater than the sum of the two unimodal conditions, a positive interaction effect was achieved as a consequence of summing cortical activations and deactivations in response to both unimodal cues. However, examination of the positive activations in both unimodal conditions show that they are not significantly different from the bimodal condition, suggesting that crossmodal inhibitory effects in sensory-specific areas during unimodal stimulation may produce artificial multisensory facilitation responses. One option, to guard against this eventuality, is to calculate interaction effects solely against positive activations only—although this is a highly conservative strategy that may underestimate the true extent of interaction sites.

Even in the event that these interactions are calculated against positive activations in the unisensory conditions, demonstrating that such effects truly reflect multisensory integration as opposed to the alternative (and likely correlated) possibility that bimodal stimuli result in increases in the perceptibility of the inputs. For example, in the case of speech, the effect of bimodal audio-visual speech may reflect enhanced comprehension when two senses are present, rather than convergence of the two sensory channels. Methods of countering such alternative interpretations have focused on manipulation of the experimental design, rather than simply analytic strategy.

3.1.3. Manipulating crossmodal congruency

Two alternative approaches to the identification of multisensory convergence zones include systematic manipulation of one or more parameters on which the integration of two modality-specific stimuli are likely to be combined (such as temporal onset and/or spatial correspondence) and more precise modeling of the features of multisensory integration observed at a neuronal level. For example, to discriminate brain areas involved in the detection of audio-visual synchrony, Bushara et al. [9] scanned subjects using PET whilst they performed a synchrony detection task. By varying the onset asynchrony between the simple auditory and visual stimuli, the authors were able to introduce different levels of task difficulty. Regression analysis was then used to identify voxels with RCBF responses that correlated positively with increasing task demand (i.e. decreasing intermodal delay). The advantage of this method is that it overcomes issues relating to the interpretation of crossmodal interaction effects because attention is balanced across the different sensory conditions.

A rather different tack was adopted by Calvert et al. [15] in which auditory and visual stimuli were matched or mismatched in terms of their presentation frequency. Brain areas putatively involved in the integration of the two inputs on the basis of their shared temporal correspondence were then identified by looking for areas showing superadditive responses to matched audio-visual stimulation compared to the sum of the two unimodal inputs, and response suppression (in which the bimodal response is significantly less than the best responding of either unimodal condition on a voxel-wise basis) in the context of mismatched inputs. The advantage of manipulating crossmodal coherence and modeling the known responses of multisensory integration at a cellular level is that the detection of opposing effects despite the same level of information in the auditory and visual channels, is less vulnerable to questions of interpretation in terms of differing levels of attention between the different conditions.

Yet another method of isolating multisensory interactions is to manipulate the perceptibility of the auditory and visual channels and model inverse effectiveness. This principle, observed at both the behavioral and electrophysiological levels, states that maximal crossmodal facilitation responses should be observed when the two unimodal stimuli are minimally effective. Callan et al. [11] adopted a similar strategy when looking for sites of integration for auditory and visual speech. By degrading the perceptibility of the auditory channel, crossmodal enhancement was found to be maximal when the seen speech was paired with auditory speech in noise, rather than in clear conditions. Thus, multisensory integration responses under these conditions are unlikely to reflect differences in the comprehensibility of speech when both channels are available (versus either modality in isolation) because the direction of the gain is orthogonal to level of perceptibility.

3.2. Electromagnetic techniques

The high temporal resolution of MEG and EEG make them ideally suited for testing hypotheses concerning the precise time course of multisensory events in the human brain. For example, neuromagnetic/electrical methods are able to answer questions about the temporal occurrence of interaction between two sensory streams (late vs early) and thus (in combination with neuroanatomical and electrophysiological data) may elucidate whether convergence arises via feedforward or feedback connections. At present, different laboratories have found evidence of first interaction effects at early (i.e. 40–46 ms

[32]), later (i.e. 120–130 ms [28,95]) and even later post-stimulus epochs (i.e. 280 ms [79]). Such discrepancies in the time course of multisensory interactions despite identical methods of analysis, suggest that stimuli and task requirements might have significant effects on the detection of integration related activity.

As with FMRI and PET, the detection of bimodally evoked responses that exceed the algebraic sum of that obtained to the two individual contributing components has also been used in the context of averaged ERP studies (e.g. [32]). In this case, amplitude values of ERP components are measured in response to stimulation in both modalities separately, as well as to concurrent bimodal stimulation. The crossmodal interaction effect is then defined as the difference waveform of [bimodal − (unimodal modality A + unimodal modality B)] at each electrode/sensor. This difference waveform can then be displayed as surface potential maps on the outer surface of the skull [27,32,69,79] or subjected to a dipole analysis to obtain an estimate of the relative strength and location of the proposed crossmodal interaction effect [27,56,95].

However, calculation of interaction effects in MEG experiments may be subject to artifacts of interpretation that may not affect haemodynamic studies to the same extent. For example, Teder-Salejarvi et al. [95] observed that if there is a component X present in all three tasks (e.g. in A, V and AV), calculating a simple interaction effect ([A + V] − AV) results in the double addition of this component in the unimodal cases, but only once in the bimodal condition ([X + X] − X = X). This effectively leads to the presence of the component X in the difference waveform and makes delineation from true multisensory components impossible. Teder-Salejarvi et al. [95] showed that an early components related to stimulus expectancy, namely the contingent negative variation (CNV, [82]), could precisely result in the appearance of a spurious early interaction effect. Similar confounding effects can be expected to result from late stimulus or task-related activity (e.g. P3 [75], motor response, etc.). To avoid these confounds, variable ISIs or a high-pass filter (after visual inspection of pre-stimulus baseline) may be applied to minimize expectancy waveforms.

The problem of unequal subtraction in simple interaction paradigms is, however, less acute in FMRI data analysis. Whereas a MEG/EEG sensor or electrode might receive signals that are generated by multiple cortical areas involved in different aspects of stimulus processing, the point of measurement in FMRI is an isolated three-dimensional unit (cubic voxels). Consequently, whereas a frontally generated CNV [82] in an MEG experiment might affect the estimation of an early auditory ERP waveform at temporal sites, similar frontal activity induced in an FMRI experiment will not affect the assessment of the BOLD signal recorded from voxels in distal regions such as the auditory cortex.

A possible alternative to the unequal subtraction problem in the neuroelectric/magnetic methods is presented in a recent EEG study by Gondan et al. [34], who used a congruency interaction paradigm to determine AV integration dynamics. With this kind of experimental design, bimodal stimuli are either congruent or incongruent on one dimension with at least two levels (e.g. spatial position). Gondan et al. [34] presented audio and visual stimuli at either of two spatial locations (S1, S2), creating congruent and incongruent audio-visual events. This resulted in the following four stimulus conditions: 1. bimodal congruent at S1; 2. bimodal congruent at S2; 3. bimodal incongruent at S1 and S2; 4. bimodal incongruent at S2 and S1. The interaction effect could then calculated as

$$\begin{aligned}&\text{congruency interaction}\\&\quad = (\text{bimodal congr. S1} + \text{bimodal congr. S2}) - (\text{bimodal incongr. S1 and S2} + \text{bimodal incongr. S2 and S1})\end{aligned}$$

With this formula, task-related activity that is present in all conditions and is not related to sensory integration (e.g. CNV, P3) is effectively canceled out; and thereby the confounding effects of unequal subtraction discussed by Hillyard and co-workers [95] (see above) are avoided and true multisensory interactions can be unambiguously accessed.

Other important information that can be extracted from the MEG signal relates to the frequency components of the measured response. The spontaneous EEG activity contains distinct rhythmic components that peak in specific frequency bands (e.g. Alpha = 8–13 Hz, Beta = 14–30 Hz; Gamma = 31–70 Hz) that correlate well with the subject's state and are modulated by stimulus characteristics and task. Using these oscillatory properties of the EEG/MEG signals, von Stein et al. [100] presented subjects with representations of the same objects through different sensory modalities (audio and visual), and were able to elucidate some of the mechanisms by which supramodal feature representation is achieved in the human brain. Coherence analysis [80] revealed an enhanced coherence between temporal and parietal electrodes in the 13–18 Hz frequency range, which was common to both modalities of presentation and absent in the control condition. This study illustrates how neuroelectrical/magnetical methods can provide unique information about large-scale oscillatory brain mechanisms involved multisensory processing and object representation. However, using similar techniques, such as SAM [88], the additional information provided by the oscillatory behavior of the EEG/MEG signals could also be potentially used to infer about the role certain oscillations play in the binding of crossmodal features and stimuli [87,97].

4. Crossmodal brain areas implicated to date

Despite the considerable variability in design and analytic strategy across different crossmodal imaging studies, several brain areas are now being consistently implicated in the multisensory synthesis of differing factors such as time, space and content. Sensitivity to shared temporal onset across different sensory cues has been shown in the superior colliculus and insula-claustrum. Several regions of the inferior and superior parietal lobe, including most explicitly, the intraparietal sulcus, appear to be involved in the detection and integration of multisensory cues based on their shared spatial location. Finally, cortex within the fundus of the superior temporal sulcus has been increasingly implicated in the integration of audio-visual speech based on the detection of shared phonetic features. In addition to these regions of heteromodal cortex, a number of recent studies now suggest that multisensory interactions also occur in early stages of the processing hierarchy, in areas of putative sensory-specific cortex [32,60].

5. Crossmodal attention and multisensory integration

A topical issue in the multisensory literature is the connection between attention and multisensory integration. This has been further sub-divided into questions concerning the role of exogenous (involuntary and stimulus-driven) and endogenous (or top-down, voluntary) attentional mechanisms on crossmodal processes (for a recent discussion see [61]).

5.1. Exogenous shifts in spatial attention or multisensory integration?

McDonald et al. [61] have argued that multisensory integration and involuntary shifts in crossmodal spatial attention may be distinct processes with separate underlying neuronal mechanisms. They base this supposition in part on behavioral data indicating that when two individual sensory cues are separated by a long time window (i.e. between 100 and 500 ms), multisensory integration and many of its perceptual consequences (e.g. ventriloquism) are greatly reduced ([44,63]). Thus, one means of discriminating multisensory integration from stimulus-driven shifts in attention is to present multisensory cues at varying degrees of asynchrony and examine the effect on the brain response (see [9]). Macaluso and Driver [58] on the other hand argue against such a clear cut distinction—citing evidence that multisensory cells such as those recorded in the superior colliculus [91] and cortex [102] still exhibit integrative responses for asynchronies extending up to 600 ms [63]. On this basis, they have argued that the distinction between crossmodal endogenous attention and multisensory integration may be simply one of terminology. An alternative but related explanation is that involuntary shifts of spatial attention may arise as a consequence of multisensory integration.

5.2. Is multisensory integration modulated by endogenous attention?

Behavioral studies have produced equivocal evidence concerning the immunity of multisensory integration to modulation by top-down voluntary attentional processes. For example, while some studies have reported effects of task *instruction* and *contextual appropriateness* on the persuasiveness of the ventriloquist's illusion, others have failed to replicate these effects (see [5] for a review of these issues). Similarly, although the direction of *spatial attention* was not found to influence the size of the ventriloquist effect [6] the same was not the case for *selective attention to modality* (Radeau [76]). Separate attempts to manipulate integration by varying *attentional load* during the perception of auditory and visual information about emotion have also provided support for the automaticity of the integrative process [101]. These data suggest that in addition to the possibility of several distinct attentional mechanisms, some but not all may influence the integration of multisensory cues. Where and under what conditions, such effects might occur, are now being addressed at the neurophysiological level.

Recently, Fort et al. [27] conducted an ERP study to examine whether the integration of auditory and visual cues pertaining to single objects was modulated depending on whether subjects directed attention towards or away from the bimodal event. Subjects were asked to discriminate two separate objects defined either by the conjunction of auditory and visual features, or by their visual or auditory features alone. In one condition, subjects were asked to attend and respond to the identity of the objects (attend). In a separate condition, they were asked to ignore the objects and respond to target items interleaved between the object trials (non-attend). Regardless of whether subjects attended or ignored the objects, bimodal interaction effects [AV > A + V] were observed in modality-specific (auditory and visual) cortices. Different crossmodal effects were found depending on the sensory dominance of the subject (i.e. whether they were visually or acoustically dominant). Beyond these early sensory areas, interaction effects in putative heteromodal regions were detected only when subjects directed their attention to the bimodal events. These findings suggest that directed attention can have a modulatory role on multisensory integration but that some features of crossmodal correspondence may be registered automatically in early sensory processing areas.

Other studies have examined whether multisensory integration is modulated by selective attention to one or other modality ([53,58]). These two studies have investigated this question in the context of spatially congruent and incongruent visual and tactile inputs. Despite the use of very similar paradigms, the two groups have obtained somewhat different results.

Using PET, Macaluso et al. [59] scanned subjects during bimanual stimulation with visual (LED) and tactile (vibration) cues. Stimulation could be single or double pulsed. There were two factors—one was covertly attended side (left or right), and the other was attended modality (vision or touch). The subjects task was to respond by saying "bip" to double pulsed cues in the attended side and attended modality—and ignore events in the unattended side and modality. When subjects were selectively attending and responding to the tactile modality, activation was detected in the post-central gyrus. Selective attention to the visual modality produced activation in the superior occipital gyrus. These modality-specific activations contrast with the response in the anterior part of the intraparietal sulcus (IPS) which was observed regardless of whether subject's attended to one or other modality, but that was specific to selective attention to modality rather than attention to side. These findings suggest that selective attention to modality enhances activation in unimodal cortex appropriate to the attended modality, and that these effects may be mediated by a supramodal attentional mechanism in the IPS.

These findings contrast to some extent with those reported from a similar study using event-related FMRI [53]. In this study, subjects were stimulated on the feet with visual, tactile or visuo-tactile cues that were presented either to the same (congruent) side or opposite sides of space (incongruent). The task was to detect and respond with a button press to either visual (attend vision condition) or tactile (attend touch condition) inputs. It was hypothesized that bimodal trials should elicit faster RTs and a superadditive brain response (VT > V + T), and this effect would be greater for spatially congruent inputs. By manipulating selective attention to modality, Lloyd et al. [53] were able to assess whether the network of areas exhibiting multisensory integration responses were subject to top-down control. The results of this study found that both behavioral and brain activation patterns differed substantially depending on whether subjects attended to the visual or tactile modality. Specifically, when attending to vision, only simultaneous tactile inputs to the same side of space speeded detection of the visual cues (see Fig. 3). Concordantly, superadditive brain responses were also only identified in the congruent bimodal condition. These were detected in a network of brain areas including the intraparietal sulcus and superior parietal lobule—areas previously implicated in processing cues in extrapersonal space—in the middle occipital and lingual gyri, consistent with the previous PET study by Macaluso and Driver [58] and in the posterior cingulate gyrus. When subjects attended to the tactile modality, however, visual cues on either side of space enhanced detection of the tactile stimuli. This suggests that under these circumstances, co-occurrence in time, rather than space, was the most salient factor leading to crossmodal facilitation of response times. Indeed, examination of the corresponding FMRI data revealed multisensory interaction effects for all bimodal presentations in the head of the caudate nucleus and claustrum—areas previously found to be sensitive to the relative temporal

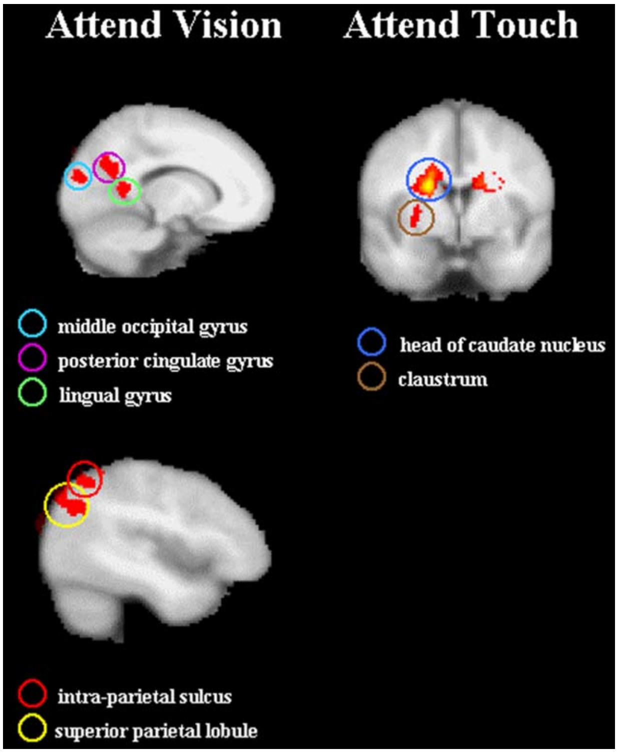

Fig. 3. The figure shows brain areas exhibiting superadditive multisensory responses when attending to the visual (left-hand side in sagittal orientation) or tactile (right-hand side in coronal orientation) modalities. Anatomical assignment of these areas (shown in red) are coded by colored circles.

onset of crossmodal cues. In sum, these data add to the growing weight of evidence suggesting that multisensory integration in humans appears to be sensitive to shifts in voluntarily directed attention and that crossmodal cues are integrated based on the most persuasive point of correspondence between them.

6. What is special about crossmodal processing?

There is now mounting evidence from a host of neurophysiological and behavioral studies that the different sensory systems interact to effect performance. But how specific are these various crossmodal phenomena to multisensory processing? Behavioral studies have shown that redundant target effects can also be observed in the context of unimodal studies. For example, it has been shown that two visual stimuli presented simultaneously and in separate hemifields can yield shorter latencies than the sum of the two unilateral stimuli (i.e. of targets presented to the contralateral or ipsilateral hemifield only) [67]. Moreover, the authors hypothesize that such synergistic effects are likely to be mediated by convergence of the contributing unisensory inputs onto populations of mutually recipient neurons. The same principle has been shown to apply to crossmodal effects mediated by multisensory neurons in the superior colliculus [91].

One apparent distinction between unisensory and multisensory interactions at both behavioral and physiological levels is the time window for integration. For example, integration of two different visual dimensions such as color and form occurs within 40 ms. However, the McGurk effect persists despite onset asynchronies of up to 180 ms. There is some evidence that response enhancement in multisensory neurons of the colliculus can be observed with asynchronies of over 600 ms. This prompts the question as to whether multisensory interactions, as distinct from unisensory ones, may be mediated by somewhat distinct mechanisms and/or involve several different collaborative networks. An alternative explanation may be that such temporal distinctions simply reflect differences in the synaptic distances required for multisensory integration compared to the integration of unimodal features within unimodal cortex. The challenge now will be to determine what to crossmodal—and the true extent of the advantages that can be gained by is specific synthesizing our sensory systems.

7. Future directions

In this review we discussed the application of haemodynamic and neuroelectical/magnetic imaging methods to the investigation of multisensory processes, and elucidated their individual strengths and weaknesses. It was emphasized that the imaging modalities discussed in this review are limited either in their temporal or spatial resolution, and by themselves do not provide sufficiently informative data about localized dynamic brain processes. However, in order to understand how the brain processes and conjoins stimulation of the different senses, high-resolution spatio-temporal imaging of brain activity is required. At present, this goal can be approximated through the integration of multiple imaging modalities [17], and by using the finding from one modality in guiding the analysis in the other. However, many basic questions regarding the relationship between the BOLD signal and neuronal activity, and thereby regarding the relationship between BOLD and electromagnetic signals remain unanswered. Shedding light on the coupling of neuronal activity with haemodynamic and large-scale network responses may make it possible to relate the spatially sensitive FMRI signal to the temporally sensitive signals of EEG and MEG.

References

[1] D.S. Barth, N. Goldberg, B. Brett, S. Di, The spatiotemporal organization of auditory, visual, and auditory–visual evoked potentials in rat cortex, Brain Res. 678 (1995) 177–190.

[2] G.C. Baylis, E.T. Rolls, C.M. Leonard, Functional subdivisions of the temporal lobe neocortex, J. Neurosci. 7 (1987) 330–342.

[3] L. Benevento, J. Fallon, B.J. Davis, M. Rezak, 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. 157 (1977) 849–872.

[4] I.H. Bernstein, M.H. Clark, B.A. Edelstein, Effects of an auditory signal on visual reaction time, J. Exp. Psychol. Hum. Percept. Perform. 80 (3, Pt. 1) (1969) 567–569.

[5] P. Bertelson, Ventriloquism: a case of cross-modal perceptual grouping, in: G. Aschersleben, T. Bachmann, J. Muesseler (Eds.), Cognitive Contributions to the Perception of Spatial and Temporal Events, Elsevier, Amsterdam, 1999, pp. 347–362.

[6] P. Bertelson, J. Vroomen, B. De Gelder, J. Driver, The ventriloquist effect does not depend on the direction of deliberate visual attention, Percept. Psychophys. 62 (2000) 321–332.

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

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

[9] K.O. Bushara, J. Grafman, M. Hallett, Neural correlates of auditory–visual stimulus onset asynchrony detection, J. Neurosci. 21 (1) (2001) 300–304.

[10] K.O. Bushara, R.A. Weeks, K. Ishii, M.J. Catalan, B. Tian, J.P. Rauschecker, M. Hallett, Modality-specific frontal and parietal areas for auditory and visual spatial localization in humans, Nat. Neurosci. 2 (1999) 759–766.

[11] D.E. Callan, A.M. Callan, C. Kroos, E. Vatikiotis Bateson, Multimodal contribution to speech perception revealed by

independent component analysis: a single-sweep EEG case study, Cog. Brain Res. 10 (2001) 349–353.

[12] G.A. Calvert, Crossmodal processing in the human brain: insights from functional neuroimaging studies, Cereb. Cortex 11 (12) (2001) 1110–1123.

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

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

[15] G.A. Calvert, P.C. Hansen, S.D. Iversen, M.J. Brammer, Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect, Neuroimage 14 (2) (2001) 427–438.

[16] G.A. Calvert, J.W. Lewis, Hemodynamic studies of audio-visual interactions, in: G.A. Calvert, C. Spence, B.E. Stein (Eds.), Handbook of Multisensory Processing, MIT Press, Cambridge, 2004.

[17] A.M. Dale, A.K. Liu, B.R. Fischl, R.L. Buckner, J.W. Belliveau, J.D. Lewine, E. Halgren, Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity, Neuron 26 (2000) 55–67.

[18] B. de Gelder, J. Vroomen, G. Pourtois, Multisensory perception of affect, its time course and its neural basis, in: G.A. Calvert, C. Spence, B.E. Stein (Eds.), Handbook of Multisensory Processing, MIT Press, Cambridge, 2004.

[19] R. Desimone, C.G. Gross, Visual areas in the temporal cortex of the macaque, Brain Res. 178 (1979) 363–380.

[20] R. Desimone, L.G. Ungerleider, Multiple visual areas in the caudal superior temporal sulcus of the macaque, J. Comp. Neurol. 248 (1986) 164–189.

[21] M.C. Doyle, R.J. Snowden, Identification of visual stimuli is improved by accompanying auditory stimuli: the role of eye movements and sound location, Percept. Psychophys. 30 (2001) 795–810.

[22] J. Driver, C. Spence, Crossmodal attention, Curr. Opin. Neurobiol. 8 (2) (1998) 245–253.

[23] J.R. Duhamel, C.L. Colby, M.E. Goldberg, Congruent representations of visual and somatosensory space in single neurons of monkey ventral intraparietal cortex (Area VIP), in: J. Paillard (Ed.), Brain and Space, Oxford University Press, New York, 1991, pp. 223–236.

[24] J.R. Duhamel, C.L. Colby, M.E. Goldberg, Ventral intraparietal area of the macaque: congruent visual and somatic response properties, J. Neurophysiol. 79 (1998) 126–136.

[25] G. Ettlinger, Object vision and spatial vision: the neuropsychological evidence for the distinction, Cortex 26 (1990) 319–341.

[26] A. Falchier, S. Clavagnier, P. Barone, H. Kennedy, Anatomical evidence of multimodal integration in primate striate cortex, J. Neurosci. 22 (2002) 5749–5759.

[27] A. Fort, C. Delpuech, J. Pernier, M.H. Giard, Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans, Cereb. Cortex 12 (2002) 1031–1039.

[28] A. Fort, C. Delpuech, J. Pernier, M.H. Giard, Early auditory–visual interactions in human cortex during nonredundant target identification, Cog. Brain Res. 14 (2002) 20–30.

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

[30] F. Frassinetti, F. Pavani, E. Ladavas, Acoustical vision of neglected stimuli: interaction among spatially converging audio-visual inputs in neglect patients, J. Cogn. Neurosci. 14 (1) (2002) 62–69.

[31] W. Fries, Cortical projections to the superior colliculus in the macaque monkey: a retrograde study using horseradish peroxidase, J. Comp. Neurol. 230 (1984) 55–76.

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

[33] S.C. Gielen, R.A. Schmidt, P.J. Van-den-Heuvel, On the nature of intersensory facilitation of reaction time, Percept. Psychophys. 34 (1983) 161–168.

[34] M. Gondan, B. Roeder, B. Niederhaus, F. Roesler, Behavior and ERP correlates of redundant visual–auditory stimulus processing, in: Proceedings of the Third Annual Meeting of the International Multisensory Research Forum, 2002, p. 80.

[35] M.S. Graziano, C.G. Gross, Spatial maps for the control of movement, Curr. Opin. Neurobiol. 8 (1998) 195–201.

[36] M.S.A. Graziano, L.A. Reiss, C.G. Gross, A neuronal representation of the location of nearby sounds, Nature 397 (1999) 428–430.

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

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

[39] R. Hari, S. Levanen, T. Raij, Timing of human cortical functions during cognition: role of MEG, Trends Cog. Sci. 4 (2000) 455–462.

[40] R. Hari, J. Rif, J. Tiihonen, M. Sams, Neuromagnetic mismatch fields to single and paired tones, Electroenceph. Clin. Neurophysiol. 82 (1992) 152–154.

[41] H. Helmhotlz, Ueber einige Gestze der Verteilung elektrischer Stroeme in koerperlichen Leitern mit Anwendung auf die tierisch-elektrischen Versuche dffd, Ann. Phys. Chem. 89 (1853) 211–233, 353–377.

[42] P. Herscovitch, J. Markham, M.E. Raichle, Brain blood flow measured with intravenous H2(15)O.I. Theory and error analysis, L. Nuc. Med. 24 (1983) 782–789.

[43] M. Hershenson, Reaction time as a measure of intersensory facilitation, J. Exp. Psych. 63 (1962) 289–293.

[44] C.E. Jack, W.R. Thurlow, Effects of degree of visual association and angle of displacement on the ventriloquism effect, Percept. Mot. Skills 37 (1973) 967–979.

[45] J.H. Kaas, C.E. Collins, The resurrection of multimodal cortex in primates: connection patterns that integrate modalities, in: G.A. Calvert, C. Spence, B.E. Stein (Eds.), Handbook of Multisensory Processing, MIT Press, Cambridge, 2004.

[46] D.C. Kadunce, J.W. Vaughan, M.T. Wallace, G. Benedek, B.E. Stein, Mechanisms of within- and cross-modality suppression in the superior colliculus, J. Neurophysiol. 78 (6) (1997) 2834–2847.

[47] K.K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, D.N. Kennedy, B.E. Hoppel, M.S. Cohen, R. Turner, H.M. Cheng, T.J. Brady, B.R. Rosen, Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 5675–5679.

[48] P.J. Laurienti, J.H. Burdette, M.T. Wallace, Y.F. Yen, A.S. Field, B.E. Stein, Deactivation of sensory-specific cortex by cross-modal stimuli, J. Cog. Neurosci. 14 (2002) 420–429.

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

[50] J.W. Lewis, M.S. Beauchamp, E.A. DeYoe, A comparison of visual and auditory motion processing in human cerebral cortex, Cereb. Cortex 10/9 (2000) 888.

[51] J.W. Lewis, D.C. Van Essen, Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey, J. Comp. Neurol. 428 (2000) 112–137.
[52] J.F. Linden, A. Grunewald, R.A. Andersen, Responses to auditory stimuli in macaque lateral intraparietal area. II. Behavioral modulation, J. Neurophysiol. 82 (1999) 343–358.
[53] D.M. Lloyd, G.A. Calvert, P.C. Hansen, X.L.F. McGlone, Visuo-tactile Integrations Sites are Modulated by Task and Attention, Society for Neuroscience, San Diego, CA, 2001.
[54] N.K. Logothetis, J. Pauls, M. Augath, T. Trinath, A. Oeltermann, Neurophysiological investigation of the basis of the fMRI signal, Nature 412 (2001) 150–157.
[55] N.E. Loveless, J. Brebner, P. Hamilton, Bisensory presentation of information, Psych. Bull. 73 (1970) 161–199.
[56] B. Lutkenhoner, C. Lammertmann, C. Simoes, R. Hari, Magnetoencephalographic correlates of audiotactile interaction, NeuroImage 15 (2002) 509–522.
[57] E. Macaluso, J. Driver, Functional imaging evidence for multisensory spatial representations and crossmodal attentional interactions in the human brain, in: G.A. Calvert, C. Spence, B.E. Stein (Eds.), Handbook of Multisensory Processing, MIT Press, Cambridge, 2004.
[58] E. Macaluso, J. Driver, Spatial attention and crossmodal interactions between vision and touch, Neuropsychologia 39 (2001) 1304–1316.
[59] E. Macaluso, C. Frith, J. Driver, Selective spatial attention in vision and touch: unimodal and multimodal mechanisms revealed by PET, J. Neurophysiol. 83 (2000) 3062–3075.
[60] E. Macaluso, C.D. Frith, J. Driver, Modulation of human visual cortex by crossmodal spatial attention, Science 289 (2000) 1206–1208.
[61] J.J. McDonald, W.A. Teder Salejarvi, L.M. Ward, Multisensory integration and crossmodal attention effects in the human brain, Science 292 (2001) 1791.
[62] H. McGurk, J. MacDonald, Hearing lips and seeing voices, Nature 264 (1976) 746–748.
[63] M.A. Meredith, J.W. Nemitz, B.E. Stein, Determinants of multisensory integration in superior colliculus neurons. I. Temporal factors, J. Neurosci. 7 (1987) 3215–3229.
[64] M. Mesulam, E.J. Mufson, Insula of the old world monkey. III: different cortical output and comments on function, J. Comp. Neurol. 212 (1982) 38–52.
[65] M.M. Mesulam, From sensation to cognition, Brain 121 (1998) 1013–1052.
[66] J. Miller, Channel interaction and the redundant-targets effect in bimodal divided attention, J. Exp. Psychol. Hum. Percept. Perform. 17 (1991) 160–169.
[67] C. Miniussi, M. Girelli, C. Marzi, Neural site of the redundant target effect electrophysiological evidence, J. Cogn. Neurosci. 10 (1998) 216–230.
[68] A.J. Mistlin, D.I. Perrett, Visual and somatosensory processing in the macaque temporal cortex: the role of 'expectation', Exp. Brain Res. 82 (1990) 437–450.
[69] S. Molholm, W. Ritter, M.M. Murray, D.C. Javitt, C.E. Schroeder, J.J. Foxe, Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study, Cognitive Brain Res. 14 (2002) 115–128.
[70] E.J. Mufson, M.M. Mesulam, Thalamic connections of the insula in the rhesus monkey and comments on the paralimbic connectivity of the medial pulvinar nucleus, J. Comp. Neurol. 227 (1984) 109–120.
[71] J.W. Neal, R.C.A. Pearson, T.P.S. Powell, The connections of area PG, 7a, with cortex in the parietal, occipital and temporal lobes of the monkey, Brain Res. 532 (1990) 249–264.
[72] S. Ogawa, R.S. Menon, D.W. Tank, S.G. Kim, H. Merkle, J.M. Ellermann, K. Ugurbil, Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model, Biophys. J. 64 (1993) 803–812.
[73] T.F. Oostendorp, J. Delbeke, D.F. Stegeman, The conductivity of the human skull: results of in vivo and in vitro measurements, IEEE Trans. Biomed. Eng. 47 (2000) 1487–1492.
[74] R. Pearson, P. Brodal, K.C. Gatter, T.P. Powell, The organization of the connections between the cortex and the claustrum in the monkey, Brain Res. 234 (1982) 435–441.
[75] J. Polich, P300 development from auditory stimuli, Psychophysiology 23 (1986) 590–597.
[76] M. Radeau, Auditory–visual spatial interaction and modularity, in: Cahiers de Psychologie Cognitive/Current Psychology of Cognition, 1994, pp. 3–51.
[77] M.E. Raichle, W.R. Martin, P. Herscovitch, M.A. Mintun, J. Markham, Brain blood flow measured with intravenous H2(15)O. II. Implementation and validation, Journal of nuclear medicine official publication, Soc. Nucl. Med. 24 (1983) 790–798.
[78] T. Raij, V. Jousmaki, MEG studies of cross-modal integration and plasticity, in: G.A. Calvert, C. Spence, B.E. Stein (Eds.), Handbook of Multisensory Processing, MIT Press, Cambridge, 2004.
[79] T. Raij, K. Uutela, R. Hari, Audiovisual integration of letters in the human brain, Neuron 28 (2000) 617–625.
[80] P. Rappelsberger, I. Szirmai, R. Vollmer, Analyse der Ausbreitung epileptischer Entladungen im Tiermodell, EEG EMG Zeitschrift fur Elektroenzephalographie, Elektromyographie und verwandte Gebiete 17 (1986) 47–54.
[81] K.S. Rockland, H. Ojima, Multimodal convergence in calcarine visual areas in macaque monkey, International Multisensory Research Forum, Geneva, Switzerland, 2002.
[82] S.K. Rosahl, R.T. Knight, Role of prefrontal cortex in generation of the contingent negative variation, Cereb. Cortex 5 (1995) 123–134.
[83] C.E. Schroeder, J.J. Foxe, The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex, Cog. Brain Res. 14 (2002) 187–198.
[84] C.E. Schroeder, R.W. Lindsley, C. Specht, A. Marcovici, J.F. Smiley, D.C. Javitt, Somatosensory input to auditory association cortex in the macaque monkey, J. Neurophysiol. 85 (2001) 1322–1327.
[85] E. Schroger, A. Widmann, Speeded responses to audiovisual signal changes result from bimodal integration, Psychophysiology 35 (1998) 755–759.
[86] R. Sekuler, A.B. Sekuler, R. Lau, Sound alters visual motion perception, Nature 385 (1997) 308.
[87] W. Singer, Synchronization of cortical activity and its putative role in information processing and learning, Annu. Rev. Physiol. 55 (1993) 349–374.
[88] K. Singh, G. Barnes, A. Hillebrand, E. Forde, A. Williams, Task-related changes in cortical synchronization are spatially coincident with the hemodynamic response, Neuroimage 16 (1) (2002) 102–114.
[89] B.E. Stein, W.S. Huneycutt, M.A. Meredith, Neurons and behavior: the same rules of multisensory integration apply, Brain Res. 448 (1988) 355–358.
[90] B.E. Stein, W. Jiang, T.R. Stanford, Multisensory integration in single neurons of the midbrain, in: G.A. Calvert, C. Spence, B.E. Stein (Eds.), Handbook of Multisensory Processing, MIT Press, Cambridge, 2004.
[91] B.E. Stein, M.A. Meredith, Merging of the Senses, MIT Press, Cambridge, 1993.
[92] B.E. Stein, M.A. Meredith, L.M. Huneycutt, L. McDade, Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli, J. Cog. Neurosci. 12 (1989) 12–24.

[93] B.E. Stein, M.A. Meredith, W.S. Huneycutt, L. McDade, Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli, J. Cog. Neurosci. 1 (1989) 12–24.

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

[95] W.A. Teder-Salejarvi, J.J. McDonald, F. Di Russo, S.A. Hillyard, An analysis of audio–visual crossmodal integration by means of event-related potential (ERP) recordings, Cog. Brain Res. 14 (2002) 106–114.

[96] J.W. Todd, Reaction to multiple stimuli, Arch. Psych. 3 (1912).

[97] A. Treisman, The binding problem, Curr. Opin. Neurobiol. 6 (1996) 171–178.

[98] B. Turner, M. Mishkin, M. Knapp, Organization of the amygdalopetal projections from modality-specific cortical association areas in the monkey, J. Comp. Neurol. 191 (1980) 515–543.

[99] R. Turner, Functional mapping of the human brain with magnetic resonance imaging, Sem. Neurosci. 7 (1995) 179–194.

[100] A. von Stein, P. Rappelsberger, J. Sarnthein, H. Petsche, Synchronization between temporal and parietal cortex during multimodal object processing in man, Cereb. Cortex. (1999) 137–150.

[101] J. Vroomen, P. Bertelson, B. de Gelder, Directing spatial attention towards the illusory location of a ventriloquized sound, Acta Psychol. (Amst) 108 (2001) 21–33.

[102] M.T. Wallace, M.A. Meredith, B.E. Stein, Integration of multiple sensory modalities in cat cortex, Exp. Brain Res. 91 (1992) 484–488.

[103] J. Watanabe, E. Iwai, Neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during visual fixation task, Brain Res. Bull. 26 (1991) 583–592.

[104] M. Watanabe, Frontal units of the monkey coding the associative significance of visual and auditory stimuli, Exp. Brain Res. 89 (1992) 233–247.

[105] R.B. Welch, D.H. Warren, Intersensory Interactions, in: K.R. Boff, L. Kaufmann, J.P. Thomas (Eds.), Handbook of Perception and Human Performance, Wiley, New York, 1986, pp. 4–5.

[106] L.K. Wilkinson, M.A. Meredith, B.E. Stein, The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior, Exp. Brain Res. 112 (1996) 1–10.

[107] D.A. Hall, I.S. Johnsrude, M.P. Haggard, A.R. Palmer, M.A. Akeroyd, A.Q. Summerfield, Spectral and temporal processing in human auditory cortex, Cereb. Cortex 12 (2002) 140–149.