

Detection of Audio-Visual Integration Sites in Humans by Application of Electrophysiological Criteria to the BOLD Effect

Gemma A. Calvert,* Peter C. Hansen,*† Susan D. Iversen,‡ and Michael J. Brammer§

*Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), John Radcliffe Hospital, Oxford OX3 1DU; †The Laboratory of Physiology, University of Oxford OX1 3PT, and ‡Department of Experimental Psychology, University of Oxford OX1 3DU, United Kingdom; and §Departments of Biostatistics and Computing, Institute of Psychiatry, London SE5 8AF, United Kingdom

Received October 16, 2000

Electrophysiological studies in nonhuman primates and other mammals have shown that sensory cues from different modalities that appear at the same time and in the same location can increase the firing rate of multisensory cells in the superior colliculus to a level exceeding that predicted by summing the responses to the unimodal inputs. In contrast, spatially disparate multisensory cues can induce a profound response depression. We have previously demonstrated using functional magnetic resonance imaging (fMRI) that similar indices of crossmodal facilitation and inhibition are detectable in human cortex when subjects listen to speech while viewing visually congruent and incongruent lip and mouth movements. Here, we have used fMRI to investigate whether similar BOLD signal changes are observable during the crossmodal integration of nonspeech auditory and visual stimuli, matched or mismatched solely on the basis of their temporal synchrony, and if so, whether these crossmodal effects occur in similar brain areas as those identified during the integration of audio-visual speech. Subjects were exposed to synchronous and asynchronous auditory (white noise bursts) and visual (B/W alternating checkerboard) stimuli and to each modality in isolation. Synchronous and asynchronous bimodal inputs produced superadditive BOLD response enhancement and response depression across a large network of polysensory areas. The most highly significant of these crossmodal gains and decrements were observed in the superior colliculi. Other regions exhibiting these crossmodal interactions included cortex within the superior temporal sulcus, intraparietal sulcus, insula, and several foci in the frontal lobe, including within the superior and ventromedial frontal gyri. These data demonstrate the efficacy of using an analytic approach informed by electrophysiology to identify multisensory integration sites in humans and suggest that the particular network of brain areas implicated in these crossmodal integrative processes are dependent on the nature of the correspondence

between the different sensory inputs (e.g. space, time, and/or form). © 2001 Academic Press

Key Words: crossmodal, audio-visual; multisensory; superior colliculus; fMRI; multiplexed designs; superadditive; interaction effects.

INTRODUCTION

The possession of multiple sensory systems provides humans and other species with clear behavioural advantages. As well as providing a framework in which information from the different modalities can be used interchangeably, the possession of several senses allows multisensory cues to be combined. Such crossmodal integration can substantially enhance our ability to detect, locate and discriminate external stimuli (for a review see Welch and Warren, 1986; Stein and Meredith, 1993). For example, psychophysical experiments have shown that reaction times to congruent inputs from more than one modality are significantly faster than those to unimodal stimuli (Miller, 1982; Hughes *et al.*, 1994; Frens *et al.*, 1995). These crossmodal advantages depend on the detection of some point of commonality between the different sensory inputs. Two major factors that determine whether different modal cues will be perceived as arising from the same event, and integrated for perceptual gain, are proximity in time and space (Radeau, 1994). Shared featural information (e.g., shape, size, intensity) can also be combined to improve object discrimination. On the other hand, asynchronous, spatially disparate, and/or semantically incongruent inputs from different modalities can be significantly less effective in eliciting responses than the unimodal stimuli alone (Stein *et al.*, 1989; Sekuler *et al.*, 1997).

Electrophysiological studies in nonhuman primates and other mammals have identified similar features of crossmodal facilitation and suppression in multisensory cells of the superior colliculus (Wallace *et al.*, 1996; Stein, 1998; King and Palmer, 1985; Jay and

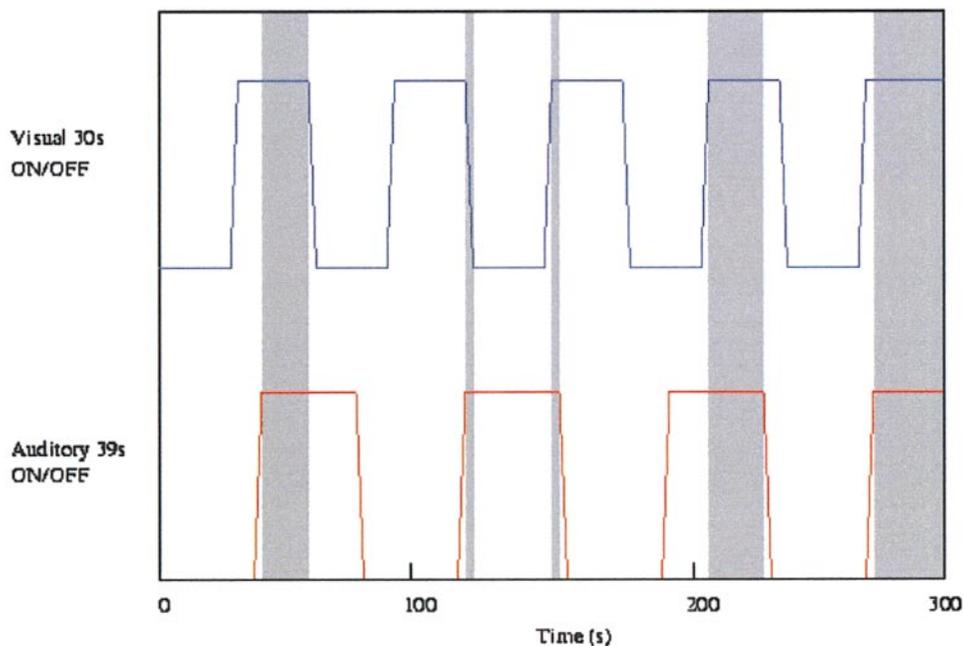


FIG. 1. Experimental paradigm: In both experiments, an overlapping multiplexed block design was used in which the visual stimulus (shown in blue) was presented every 30 s and the auditory stimulus (shown in red) was turned on every 39 s. Both of these modalities alternated with a resting condition. This design allowed subjectively unpredictable periods of auditory, visual, audio-visual, and rest conditions to be presented over the total scan time of 5 min per experiment (time in seconds is shown on the *x* axis). Grey bars indicate the periods of coincident auditory and visual stimulation. Figure reprinted from *Current Biology* (Calvert *et al.*, 2000) with permission from Elsevier Science.

Sparks, 1984), a midbrain structure involved in attention and orientation to external stimuli. Each multisensory cell in the superior colliculus contains a map of sensory space, one for each modality to which it responds (King and Palmer, 1985; Meredith and Stein, 1986). When sensory cues from more than one modality occur simultaneously and arise from the same spatial location, the firing rate of these cells is substantially enhanced, often to a level greater than that predicted by summing responses to the individual components (Wallace *et al.*, 1996; Meredith and Stein, 1986). These crossmodal enhancements of neuronal activity are maximal when the contributing unimodal stimuli are minimally effective—a principle referred to as inverse effectiveness (Meredith and Stein, 1986). In contrast, asynchronous and/or spatially disparate cues can produce profound response depression (Kadunce *et al.*, 1997). In this case, a vigorous response to a unimodal stimulus may be substantially inhibited or even eliminated by the concurrent presentation of a conflicting stimulus from another modality. These features of multisensory integration, and the rules which determine binding, have also been shown to apply to the orienting and attentive behaviors mediated by the superior colliculus (Stein *et al.*, 1989; Wilkinson *et al.*, 1996). Similar principles of intersensory synthesis have also been shown to operate at the cortical level (Wallace *et al.*, 1992; Wilkinson *et al.*, 1996). However,

slight differences in spatial sensitivity between these cells and those studied in the colliculus implies that the integration of spatial and non-spatial (i.e., identity) information may be supported by different populations of multisensory cells distributed in geographically distinct locations throughout the brain (Mesulam, 1998).

These parallels between the response properties of multisensory cells, so far only studied in nonhuman animals, and the effects of crossmodal processing on human behavior are provocative, suggesting that multisensory integration in humans is achieved via a similar mechanism as that identified in the superior colliculus and cortex in other species and that response enhancement and depression may be hallmark features of these intersensory interactions. Evidence supporting this hypothesis was recently obtained using functional magnetic resonance imaging (fMRI) in a study of audio-visual speech perception (Calvert *et al.*, 2000). By instructing subjects to listen to speech while viewing congruent or incongruent lip and mouth movements and using the BOLD response as an indicator of neuronal activity, we detected signal changes in the left superior temporal sulcus (STS) during the crossmodal integration of “what” information that resemble the responses of multisensory cells to spatially congruent and incongruent audio-visual stimuli. Specifically, congruent audio-visual speech resulted in a superadditive response enhancement in the STS (compared to

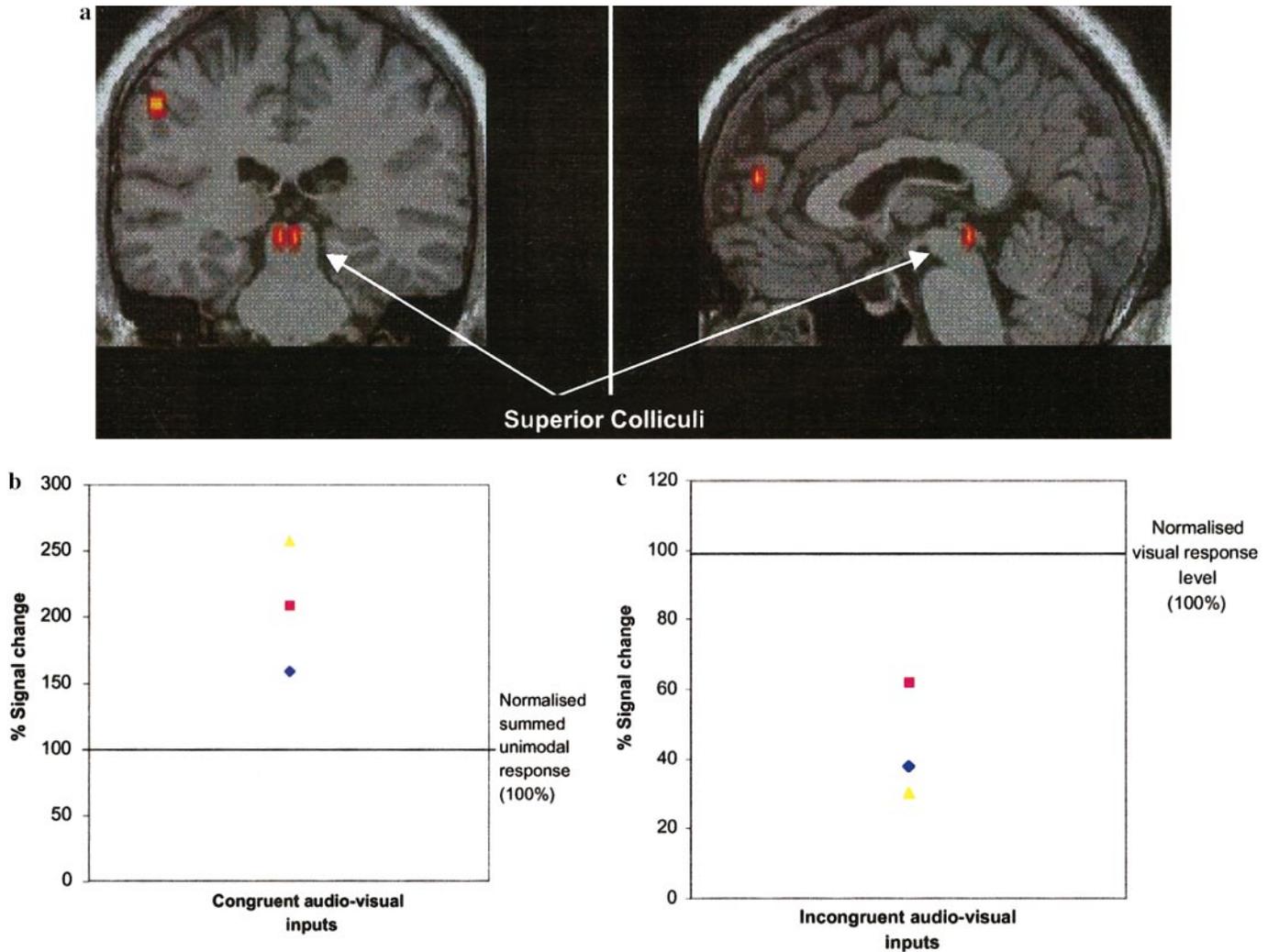


FIG. 2. (a) The focus of activity in the superior colliculi (indicated by the white arrows) in the group map is shown in coronal (left panel) and sagittal (right panel) orientations. (b) The panel illustrates the size of the superadditive signal changes in each of the 3 voxels comprising the cluster of activations observed in the superior colliculus illustrated in a. The extent of superadditivity for each voxel is shown by comparison against the normalised and summed unimodal responses detected in this cluster. (c) The panel illustrates the size of the response depression in each of these 3 voxels as a percentage of the largest (visual) unimodal response.

the sum of responses obtained when subjects were exposed to either audible or visual speech independently) and incongruent audio-visual signals produced response depression. In this case, the BOLD response to bimodal speech was significantly lower than the summed response to the two unimodal conditions. These data thus provided preliminary evidence that multisensory cells have a functional role in integrating information across modalities in humans and that this mechanism may extend to the crossmodal integration of “what” as well as “where” information.

Here, we describe an fMRI study designed to investigate whether similar crossmodal gains and decrements in the BOLD response can be observed during the integration of nonspeech auditory and visual inputs, matched or mismatched solely on the basis of their temporal synchrony. Synchrony has been shown

to be a major determinant of intersensory binding at both the neuronal and behavioural levels and can even be detected by infants as young as two months old (Lewkowicz, 1996). By using an identical fMRI paradigm to that used to identify sites of integration during audio-visual speech, we were also able to determine whether similar brain areas are involved in the synthesis of speech and nonspeech inputs.

MATERIALS AND METHODS

Subjects

Ten right-handed native English speaking subjects (mean age 30.1, range 22–45 years: five males and five females) participated in the study. All subjects were in good health with no past history of psychiatric or neu-

rological diseases and gave informed consent to the protocol which had been approved by the local Research Ethics Committee. Subjects had normal or corrected-to-normal (with contact lenses) visual acuity.

Experimental Design

Subjects were scanned during two 5-min experiments (synchronous and asynchronous conditions) in which visual and auditory stimuli were presented within a multiplexed block design (Fig. 1). This paradigm allows two on-off box-car designs to be presented within the same run at different alternation rates. Thus, the visual stimulus alternated with a black screen every 30 s. The auditory stimulus alternated with a "silent" resting condition every 39 s. This design produced subjectively unpredictable periods of rest, auditory stimulation alone, visual stimulation alone, and periods when the auditory and visual stimuli occurred at the same time. Approximately equal numbers of brain volumes were acquired in each condition. Visual stimuli were back-projected onto a screen at the end of the scanner bed. Auditory stimuli were presented from the audio output of a video recorder via a pneumatic headset designed to minimise interference from scanner noise. Prior to running the experiments, all subjects completed a routine check to ensure they could hear the auditory stimuli over the scanner noise.

In both experiments, the visual stimulus consisted of a reversing checkerboard pattern of alternating black and white squares (each square subtending $0.4^\circ \times 0.4^\circ$) in a 20×16 -inch grid. This grid subtended a visual angle of $13^\circ \times 13^\circ$, almost filling the subject's entire field of view down the bore of the magnet. The squares changed colour from black to white or vice versa at an alternation frequency of 8 Hz (125 ms between reversals). Every 30 s the reversing checkerboard pattern itself alternated with a uniform grey field of roughly the same mean luminance. The audible stimuli comprised 100 ms white noise bursts. In the first (synchronised audio-visual) experiment, the white noise bursts repeated at a rate of 8 Hz (i.e., 100 ms on with a 25-ms gap), so that in periods when the auditory and visual stimuli were coincident, the onset of each noise burst coincided with the onset of the reversal of the checkerboard. In the second (asynchronous audio-visual) experiment, the reversal rate of the visual checkerboard remained the same (8 Hz), but the onset of the white noise bursts were randomly phase-shifted with respect to the onset of the reversal of the checkerboard. Thus, when the auditory and visual stimuli were on together, they were obviously out of synchrony. The periods of coincident auditory and visual stimulation ranged from 3–27 s. The order of presentation of these two experiments was randomized across subjects.

To avoid activation due to response selection from confounding activations relating to the integration and

separation of congruent and incongruent inputs respectively, subjects were not required to make any sort of response during the experiment. Instead, they were instructed to concentrate on the projection screen located at the end of the scanner bed for the duration of the experiment and to passively attend to the auditory and visual modalities equally.

To detect the presence of a multisensory binding site using the BOLD response, stricter criteria need to be imposed than that stipulated by those studying single cells. Although simple linear summation may be sufficient to illustrate multisensory integration in single cells, it is an insufficient criterion for fMRI. This is because the BOLD signal represents the averaged response from a large population of cells within the same area. Simple linear summation of effects could thus indicate stimulation of separate populations of unimodal auditory and visual cells within a single voxel sampled in fMRI. In the current study we have thus chosen to use interaction effects to infer crossmodal facilitation and inhibition. Areas meeting criteria for a site of crossmodal integration would therefore need to show a positive interaction effect in the presence of congruent audio-visual inputs (where $[av > (a + v)]$) and a response depression in the presence of incongruent audio-visual inputs $[av < \max(a, v)]$, where $\max(a, v)$ indicates the greater of a or v alone. In our previous study of audio-visual speech perception (Calvert *et al.*, 2000), we used a negative interaction effect that was subadditive $[av < (a + v)]$. Response depression is by definition subadditive, but subadditive interactions need not be response depressions. In this study we have used response depression, rather than subadditivity, because it adheres more closely to the electrophysiological criteria for crossmodal inhibition, and unlike a subadditive interaction, protects against a potential explanation in terms of saturation-like behavior in the nonlinear translation of neural activity to hemodynamic responses. By requiring the response to both modalities to be less than the larger of the unimodal responses, this ensures we are looking at true negative interactions at the neural level.

Image Acquisition

Gradient-echo echo-planar (EPI) MR images were acquired using a 1.5 Tesla GE Signa system retrofitted with Advanced NMR operating console with a quadrature birdcage control. One hundred T2*-weighted images depicting BOLD contrast (Ogawa *et al.*, 1990) were acquired over 5 min at each of 14 near-axial noncontiguous 7-mm thick planes parallel to the intercommisural (AC-PC) line: TE = 40 ms, TR = 3 s, in-plane resolution 3 mm, interslice gap = 0.7 mm. An inversion recovery EPI dataset was also acquired at 43 near-axial 3-mm planes parallel to the AC-PC line to facilitate registration of fMRI datasets to the standard

TABLE 1

| | Talairach coordinates | | | Cluster size | P value | Side | B.A. |
|---|-----------------------|-----|-----|--------------|-----------|------|-------|
| | x | y | z | | | | |
| A. Auditory only—matched experiment | | | | | | | |
| Superior temporal gyrus | -50 | -26 | 11 | 158 | ≤0.000005 | L | 41/42 |
| Superior temporal gyrus | 57 | -21 | 8 | 115 | ≤0.000005 | R | 42/22 |
| Medial frontal pole | -12 | 52 | -3 | 44 | ≤0.000005 | L | 10 |
| Middle temporal gyrus | 56 | -16 | -10 | 7 | 0.000075 | R | 21 |
| | -39 | -64 | 20 | 5 | 0.00006 | L | 22/39 |
| Insula | 49 | 19 | -2 | 7 | 0.00006 | R | — |
| Inferior frontal gyrus | 55 | 6 | 26 | 5 | 0.00006 | R | 44/6 |
| B. Auditory only—mismatched experiment | | | | | | | |
| Superior temporal gyrus | 53 | -14 | 10 | 181 | 0.000018 | R | 41/42 |
| | -51 | -28 | 14 | 94 | 0.000009 | L | 42 |
| Posterior cingulate gyrus | 1 | -44 | 11 | 35 | ≤0.000005 | M | 29/30 |
| Middle frontal gyrus | 34 | 17 | 48 | 22 | ≤0.000005 | R | 6 |
| Medial frontal pole | 2 | 51 | -15 | 15 | 0.000018 | M | 11 |
| Insula | -29 | 26 | -10 | 13 | 0.000129 | L | — |
| Inferior parietal lobule | 47 | -25 | 42 | 12 | 0.000319 | R | 40 |
| | -36 | -64 | 37 | 10 | 0.000324 | L | 40 |
| Middle temporal gyrus | 48 | 3 | -13 | 7 | 0.000171 | R | 21 |
| C. Visual only—matched experiment | | | | | | | |
| Lingual gyrus | 1 | -74 | 0 | 1812 | ≤0.000005 | M | 17/18 |
| Precuneus | 22 | -61 | 40 | 144 | ≤0.000005 | R | 7 |
| | -10 | -70 | 38 | 44 | ≤0.000005 | L | 7 |
| | -13 | -56 | 48 | 21 | ≤0.000005 | L | 7 |
| Posterior cingulate gyrus | -27 | -46 | 41 | 13 | 0.00008 | L | 23 |
| Superior temporal sulcus | 56 | -31 | 4 | 11 | ≤0.000005 | R | 21/22 |
| Inferior parietal lobule | -35 | -44 | 48 | 4 | ≤0.000005 | L | 40 |
| D. Visual only—mismatched experiment | | | | | | | |
| Lingual gyrus | 5 | -80 | 3 | 1874 | ≤0.000005 | R | 17/18 |
| Inferior frontal gyrus | 47 | 4 | 32 | 226 | ≤0.000005 | R | 44/45 |
| | -45 | 3 | 27 | 150 | ≤0.000005 | L | 44/45 |
| Precuneus | -17 | -64 | 43 | 83 | ≤0.000005 | L | 7 |
| Insula | -38 | 13 | -5 | 52 | 0.00007 | L | — |
| | -42 | 28 | 7 | 15 | 0.00006 | L | — |
| | 45 | -16 | -4 | 14 | 0.00006 | R | — |
| Inferior parietal lobule | -35 | -45 | 51 | 35 | 0.0006 | L | 40 |
| Superior temporal sulcus | -53 | -30 | 9 | 7 | 0.00009 | L | 21/22 |
| Superior colliculus | 0 | -32 | 2 | 6 | 0.00006 | M | — |

Note. L, left hemisphere; R, right hemisphere; M, midline. The Talairach coordinates represent the peak activation within the cluster.

stereotactic space (Talairach and Tournoux, 1988) [TE = 80 ms, TI = 180 ms, TR = 16 s, in-plane resolution 3 mm].

Data Analysis

Prior to time-series analysis, the data for each subject were pre-processed to remove low-frequency signal changes and minimise movement-related artefacts. Linear modelling of the multiplexed audio-visual data at each voxel was then used to quantify unimodal responses and interaction effects.

A standard GLM model was fitted:

$$Y_t = m + v.CV_t + a.CA_t + av.CAV_t + \varepsilon_t$$

where y is the image intensity at time point t and CV ,

CA , CAV are the convolutions of the epochs of auditory alone (A), visual alone (V), and simultaneous auditory-visual (AV) stimulation with a Poisson function simulating a haemodynamic delay of 6 s. The amplitudes of the responses to auditory, visual and audio-visual stimulation are given by (a), (v), and (av) and the mean image intensity by (m). The residual error at time t following removal of residual autocorrelation is ε_t . The values of a , v , and av computed in the above model represent deviations of the image intensity during the A , V , and AV epochs from its overall mean value (m) across the whole time series. The aim of the analysis was to test the significance of effects analogous to superadditivity and response depression identified in electrophysiological studies (e.g., Stein and Meredith, 1993). These effects can be written as [$av > (a + v)$] for

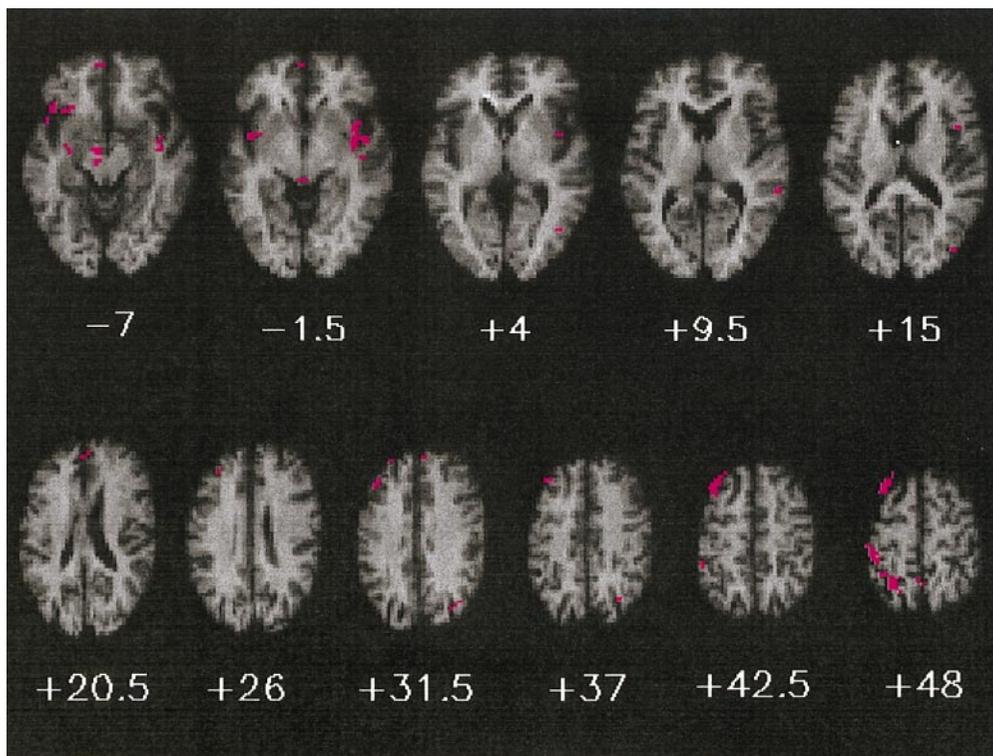


FIG. 3. The figure shows the group activation map of the network of brain areas exhibiting both superadditive response enhancements and response depression in response to synchronous and asynchronous audio-visual inputs. This network includes insula, superior colliculus, superior temporal sulcus, intraparietal sulcus, and regions of medial ventral and dorsal frontal cortex. The location of the slices in the z plane are indicated below each slice. Activations are shown in radiological convention and coloured purple.

superadditivity and $[av < \max(a, v)]$ for response depression.

Significant unimodal effects, superadditive effects or response depressions were identified at each voxel by randomization techniques as described previously (Calvert *et al.*, 2000) both at the individual subject level and group level. Briefly, the basic methodology involved fitting the above model and estimating the values of the unimodal responses, the superadditive effect and response depression at each voxel. The same estimations were then made after randomisation of the fMRI time-series ten times at each voxel. The randomization procedure preserves the distribution of image intensity fluctuations in the time series but renders them unrelated to the experiment, thus achieving the null hypothesis of no experimentally determined effect (see Bullmore *et al.*, 1996). The “randomized” estimates of a , v , superadditivity and response depression were then pooled over all voxels to create a distribution of each response of interest under the null hypothesis. The probability of any value of a unimodal response (a , v), a superadditive effect or a response depression under the null hypothesis could then be determined by sampling each null distribution at the appropriate point. For example, a one-tailed test at $P = 0.001$ is carried out by sorting the data in each distribution by magnitude and determining the value of the sorted

parameter, which is only exceeded by 0.1% of all the values in the distribution. This value is then the critical value for testing at $P = 0.001$ and any value of the observed parameter exceeding that critical value has a probability under the null hypothesis of 0.001 or less.

For group analysis, the observed and randomized statistical data were transformed into the stereotactic space of Talairach and Tournoux (1988). Overlap maps between the interactions observed in each experiment were made after thresholding each experimental group map to identify voxels with a Type 1 error probability of 0.001. At this P value, with approximately 20,000 intracerebral voxels remapped into 25 slices following Talairach transformation, we would expect only 20 type 1 errors in the whole activation map, i.e., <1 voxel per slice of the Talairach template.

Superimposition maps of unimodal effects were also constructed by identifying areas responding to both unimodal stimuli, i.e., maps where the condition (a and v) was obeyed after thresholding both unimodal maps at a type I error probability of 0.001.

RESULTS

The group results obtained during each experiment and across both experiments were as follows:

Auditory-Only Epochs

Across both experiments, passive listening to white noise bursts in the absence of the visual checkerboard produced predictable bilateral activation of primary (Brodmann Area (B.A.) 41), secondary (B.A. 42), and association auditory cortices (B.A. 22/21; Tables 1A and 1B). This activation extended into the depths of the superior temporal sulci and spread posteriorly onto the ventral bank of the middle temporal gyrus (B.A. 21/22). In the first (audio-visual matched) experiment (Table 1A) other areas activated during the auditory only epochs included a large cluster of (44) activated voxels in the left medial frontal pole (B.A. 10), the superior portion of the left posterior middle temporal gyrus extending into the inferior parietal lobule (B.A. 39), the right insula cortex and the superior tip of the right inferior frontal gyrus (B.A. 44/6). In the second (audio-visual mismatched) experiment (Table 1B) additional areas activated included the posterior cingulate gyrus (B.A. 29/30), right middle frontal gyrus (B.A. 6), left insula and the intra-parietal sulcus (B.A. 40) bilaterally.

Visual-Only Epochs

Across both experiments, passive viewing of a rapidly alternating black and white checkerboard resulted in large areas of activation in the occipital lobe extending from the location of the peak activation in V1 (primary visual cortex) to extrastriate regions, anterior and laterally into the occipito-temporal junction incorporating parts of the V5 complex and associated extrastriate areas (MT+) (Watson *et al.*, 1993) and superiorly into the precuneus (B.A. 7; Table 1C and 1D). Activation was also observed in the left inferior parietal lobule (B.A. 40). In the matched audio-visual experiment, additional areas activated in the visual only periods included the posterior cingulate (B.A. 23) and right STS (B.A. 21/22). Visual only periods in the mismatched audio-visual experiment stimulated bilaterally the inferior frontal gyrus (B.A. 44/45) and insula, the left STS (B.A. 21/22) and the superior colliculus.

Superimposition of Unimodal Auditory and Visual Maps

One strategy that has been used to infer sites of multisensory integration in recent years has been to overlap the activation maps derived from two separate experiments involving different modalities and to identify sites of coincident activation (e.g., Calvert *et al.*, 1997; Bushara *et al.*, 2001; Macaluso *et al.*, 2000a). To compare this strategy with the use of interaction effects, we performed the same analysis on the auditory and visual activation maps derived from the unimodal conditions obtained in both the temporally matched and mismatched experiments. Notwithstanding the

fact that the unimodal maps were thresholded at a level that produced extensive activation for these stimuli, superimposition of the combined auditory-only onto the combined visual-only maps revealed no significant regions of coincident activation.

Superadditivity and Response Depression to Synchronous and Asynchronous Bimodal Stimuli

The most highly significant superadditive response enhancement and corresponding response depression in the presence of temporally matched and mismatched audio-visual stimuli was detected in the superior colliculi (Fig. 2; Table 2). It is noteworthy that despite the potential for MR signal loss in the superior colliculi due to pulsation artefacts, we were nevertheless able to identify responses in these structures by modelling interaction effects and therefore did not need to apply a region of interest approach. However, the superior colliculi were not alone in exhibiting both significant superadditive effects and response depression. Rather, they formed part of a network of brain areas (Fig. 3) displaying behavior reminiscent of that observed in electrophysiological studies of multisensory cells. This network included right inferior frontal gyrus (B.A. 8), multiple sites within the right lateral sulcus (B.A. 38/47), and ventromedial frontal gyrus (B.A. 10). The crossmodal effects in these frontal cortical sites were widely distributed and confined to the right hemisphere. Significant response enhancement and depression was also observed in the insula bilaterally (left more extensively than right), in the right superior parietal lobule and right inferior parietal sulcus (B.A. 7/40), in the left superior occipital gyrus and left superior temporal sulcus (B.A. 22/21).

Superadditive Effects to Synchronized Audio-Visual Stimuli Only

Other brain areas showed superadditive effects to synchronized audio-visual stimuli but failed to exhibit a correspondingly significant response depression to asynchronous inputs. The most highly significant superadditive interaction effects were observed in inferior and superior regions of the left parietal lobe (B.A. 39/40). Other areas included middle frontal gyrus bilaterally (B.A. 6/8/9), the anterior and posterior cingulate gyri (B.A. 32/30), the lingual gyrus (B.A. 17/18), right precentral gyrus (B.A. 4), left middle temporal gyrus (B.A. 21), and the left claustrum.

Response Depression to Asynchronous Audio-Visual Stimuli Only

There were also several brain areas that displayed response depression to asynchronous audio-visual inputs but did not exhibit positive interaction (i.e., superadditive) effects to synchronised audio-visual stim-

TABLE 2
Superadditive and Response Depression

| | Talairach coordinates | | | Cluster size | P value | Side | B.A. |
|----------------------------------|-----------------------|-----|-----|--------------|----------|------|-------|
| | x | y | z | | | | |
| Superior colliculus | 3 | -28 | -2 | 3 | 0.000014 | M | — |
| Insula | 40 | 4 | 0 | 8 | 0.000328 | R | — |
| | -40 | 4 | -1 | 41 | 0.00171 | L | — |
| Frontal regions | | | | | | | |
| Superior frontal gyrus | 36 | 24 | 40 | 47 | 0.000058 | R | 8 |
| Ventromedial frontal gyrus | 3 | 56 | -4 | 8 | 0.001255 | R | 10 |
| | 3 | 45 | 20 | 8 | 0.003081 | R | 10 |
| Anterio-posterior lateral sulcus | 41 | 25 | -7 | 8 | 0.00002 | R | 47 |
| | 43 | 13 | -7 | 5 | 0.00002 | R | 38 |
| | 31 | 23 | -7 | 7 | 0.000051 | R | 47 |
| | 35 | 2 | -13 | 5 | 0.000187 | R | 38/47 |
| Superior temporal sulcus | -51 | -36 | 9 | 6 | 0.000701 | L | 21/22 |
| Parietal Regions | | | | | | | |
| Inferior parietal sulcus | 47 | -32 | 48 | 15 | 0.000109 | R | 40 |
| Superior parietal lobule | 29 | -55 | 50 | 28 | 0.000133 | R | 7 |
| Superior occipital gyrus | -28 | -74 | 30 | 10 | 0.000207 | L | 19 |
| | -25 | -67 | 37 | 6 | 0.000894 | L | 19 |

uli. These included the right inferior frontal (B.A. 44/45) and middle temporal (B.A. 21) gyri, the right cerebellum, the left precentral sulcus (B.A. 6) and left hippocampus (B.A. 28/36), and the fusiform gyri (B.A. 18) bilaterally.

DISCUSSION

This study has demonstrated that electrophysiological characteristics of multisensory integration at the neuronal level can also be detected in the BOLD re-

sponse in human fMRI studies. Specifically, the marked response enhancements and depressions observed in multisensory cells in response to spatially congruent and incongruent bimodal cues are paralleled by superadditive bimodal gains and corresponding response decrements in the BOLD signal during the perception of synchronous and asynchronous audio-visual stimuli. The network of areas exhibiting this behavior is distinct from that identified during our previous study of audio-visual speech (Calvert *et al.*, 2000) but resembles closely that reported in another very recent

TABLE 3

| | x | y | z | Cluster size | P value | Side | B.A. |
|-------------------------------|-----|-----|-----|--------------|-----------|------|-------|
| A. Superadditive effects only | | | | | | | |
| Parieto-occipital junction | -35 | -61 | 31 | 67 | 0.00008 | L | 39/40 |
| Inferior parietal lobule | -35 | -24 | 26 | 11 | 0.00047 | L | 40 |
| Middle frontal gyrus | 37 | 22 | 42 | 43 | 0.00007 | R | 9 |
| | -22 | 24 | 46 | 32 | 0.00008 | L | 6/8 |
| Middle temporal gyrus | -58 | -11 | -6 | 14 | 0.00009 | L | 21 |
| Lingual gyrus | -3 | -87 | -2 | 17 | 0.00008 | M | 17/18 |
| Anterior cingulate gyrus | 11 | 26 | 19 | 17 | 0.00014 | R | 32 |
| Posterior cingulate gyrus | 1 | -46 | 15 | 14 | 0.00017 | M | 30 |
| Precentral gyrus | 45 | -6 | 48 | 8 | 0.00009 | R | 4 |
| Claustrum | -20 | -33 | 4 | 7 | 0.00031 | L | — |
| B. Response depression only | | | | | | | |
| Inferior frontal gyrus | 48 | 7 | 20 | 47 | ≤0.000005 | R | 44/45 |
| Middle temporal gyrus | 48 | -50 | -2 | 42 | ≤0.000005 | R | 21 |
| Fusiform gyrus | 15 | -80 | -14 | 40 | ≤0.000005 | R | 18 |
| | -14 | -78 | -13 | 32 | ≤0.000005 | L | 18 |
| Hippocampal gyrus | -18 | -21 | -18 | 18 | ≤0.000005 | L | 28/36 |
| Cerebellum | 33 | -47 | -18 | 10 | ≤0.000005 | R | — |
| Precentral sulcus | -45 | -3 | 42 | 6 | 0.000023 | L | 6 |

study of audio-visual synchrony detection using positron emission tomography (PET) (Bushara *et al.*, 2001). Similar crossmodal gains and decrements have now also been found in other brain areas using a range of different human imaging techniques and combinations of modalities (e.g., Giard *et al.*, 1999; Lewis *et al.*, 2000; Macaluso *et al.*, 2000b; Foxe *et al.*, 2000). Taken together with these data, our findings add to the growing weight of evidence implicating the involvement of multisensory neurons in the crossmodal synthesis of sensory information in humans. They also suggest that the particular network of brain areas involved in this process is dependent on the nature of the stimuli and the bases (e.g., temporal, spatial, and/or content-related congruity) on which they are combined.

Of the brain areas exhibiting superadditive response enhancements and response depression to synchronous and asynchronous audio-visual inputs, the most highly significant combined effects were observed in the superior colliculi. Given the emphasis in the present study on crossmodal synchrony, the detection of this structure was unexpected. Although synchrony is one important determinant of crossmodal integration in multisensory collicular cells, it is a seemingly less critical influence than spatial coherence (Meredith *et al.*, 1996). Indeed, multisensory cells in the superior colliculus have been shown to integrate multisensory cues despite quite long temporal displacements between the unimodal stimuli (Wallace *et al.*, 1992), providing these inputs are in close spatial alignment. However, presentation of the visual stimuli on a screen located at the end of the scanner bed and the auditory information via headphones also precludes such spatial correspondence. One explanation for the unexpected detection of superior colliculi activation in the current study is afforded by behavioural studies of crossmodal links in attention (see Driver and Spence, 1998). These studies have shown that when two or more sensory inputs are in close temporal proximity, albeit in slightly disparate spatial locations, they are generally perceived as emanating from a common event (e.g., Radeau, 1994; Driver *et al.*, 1996). Thus, although our stimuli were not technically in the same spatial location, when the onset of the auditory cues were time-locked to the reversal of the checkerboard, subjects reported perceiving the sounds as emanating from the location of the visible stimulus. This crossmodal illusion is consistent with the synthesis of the auditory and visual coordinate information in multisensory cells in the superior colliculus.

Despite the physiological plausibility of superior colliculi involvement in this study, it is worth noting that multisensory cells (at least in those species studied) are only a subset of those contained in this structure. As our visual stimulus subtended only a portion of central visual space, only those neurons in the rostral

superior colliculus might have been expected to show integration responses. However, information pertaining to the activity within subsections of the superior colliculus is beyond the resolution of the voxel-size used in the current study. Nevertheless, the fact that we have detected significant superadditive effects and response depression in three voxels (approx 135 mm³) overlying approximately the whole extent of the superior colliculi suggests that the effects occurring in the integrative cells must be of considerable magnitude to translate to detectable BOLD interaction effects.

Significant, but weaker, bimodal response enhancements and response decrements were also detected in a number of cortical regions. These included the left superior occipital and superior temporal sulci, the right inferior parietal sulcus and superior parietal lobule, the right ventromedial and superior frontal gyri and the insula bilaterally. Anatomical studies in monkeys have shown that, with the exception of the superior occipital gyrus, all these areas receive afferent connections from more than one modality or are monosynaptically connected to multisensory association areas (Jones and Powell, 1970; Pandya and Yeterian, 1985; Seltzer and Pandya, 1989; Mesulam and Mufson, 1982). Furthermore, electrophysiological studies have shown that these regions contain cells responsive to stimulation in more than one sensory modality (Hikosaka *et al.*, 1988; Bruce *et al.*, 1981; Andersen *et al.*, 1997; Colby and Duhamel, 1991; Watanabe and Iwai, 1991). Our findings, together with these data from animal studies, thus support a model of crossmodal synthesis within a distributed network of brain areas. The idea of interactive communication between the superior colliculus and a network of cortical regions gains support from recent electrophysiological studies in cats. These investigations have indicated that the crossmodal integrative functions of the superior colliculus depend on top-down influences from areas of association cortex. For example, temporary paralysis of the anterior ectosylvian fissure and rostral lateral sulcus by cryogenic deactivation eliminates the characteristic crossmodal response enhancement in multi-sensory integration cells in the colliculus (Wallace and Stein, 2000). These data are particularly intriguing in light of the current study which found that the most highly significant interaction effects were found in the superior colliculus and lateral sulcus, respectively.

It seems unlikely however that all the regions identified in the study are involved in the initial integrative event. Rather, there is growing support for the suggestion that while some areas participate in the initial integration of sensory information (e.g., Stein, 1998; Calvert *et al.*, 2000), the response in other areas may be modulated as a downstream consequence of this synthesis (e.g., Calvert *et al.*, 1999; Macaluso *et al.*, 2000b). Different brain areas may also have a circum-

scribed involvement in the crossmodal integration of different types of sensory information. For example, Bushara *et al.* (2001) found that the highest task-related activity during audio-visual synchrony detection was identified in the insula, suggesting a particular role for this region in mediating temporally defined crossmodal interactions. Indeed, significant superadditive enhancement and response depression was also detected in the insula in the current study which involved presentation of synchronous and asynchronous audio-visual inputs. The intraparietal sulcus and superior parietal lobule have been previously implicated in human imaging experiments investigating the influence of spatial concordance on the integration of audio-visual and visuo-tactile inputs (Eimer and Schroger, 1998; Eimer, 1999; Macaluso *et al.*, 2000a, 2000b). Finally, the identification of significant crossmodal interaction effects during the integration of heard speech with lip-read (Calvert *et al.*, 2000) or written (Raij *et al.*, 2000) linguistic material suggests that polysensory regions within the fundus of the STS are particularly involved in the intersensory synthesis of identity information.

The current study has demonstrated the utility of exploiting interaction effects to identify putative sites of multisensory integration. The superior colliculus was not detectable by Bushara *et al.* (2001), using simple subtraction methods of data analysis, and was only identified following post hoc correlational analysis. We also failed to identify this structure using simple superimposition of the activation maps from auditory and visual stimulation alone. However, detection of a possible integration site by superimposition relies on the presence of a detectable BOLD response to each unimodal condition. Electrophysiological studies on the other hand have shown that not all cells that exhibit bimodal response enhancement necessarily exhibit suprathreshold responses to each unimodal component (Stein and Meredith, 1993). The detection of superadditive interaction effects by our analytic methods is not subject to such a constraint, though response depression requires a significant response to at least one unimodal stimulus. It should be noted, however, that our requirement for an area to exhibit both superadditivity and response depression (based on direct analogy with single-cell electrophysiological findings) is a somewhat strict criterion for the identification of multisensory integration sites. We have found that not all areas exhibiting superadditivity necessarily also show response depression and vice versa. However, when thresholded at the same level ($P \leq 0.001$), the brain regions meeting either one, but not both of these criteria are very similar (superior colliculus, STS, inferior parietal lobe, insula, and medial frontal lobe). This may indicate that essentially the same networks are sampled by the two criteria, many of the differences reflecting the fact that thresholding the activa-

tion maps produces effects dependent on the relative local levels of superadditivity and response depression. There are, however, some differences that are worthy of note. For example, areas showing superadditive effects to concordant stimuli but no significant response depression with discordant inputs include the lingual gyrus, posterior cingulate gyrus and precuneus. Response depressions, with no parallel superadditive effects, were found in the hippocampus and fusiform gyrus. The former may in part reflect attentional effects (Davis *et al.*, 2000; Corbetta *et al.*, 1991; Macaluso *et al.*, 2000a), the latter the demands of learning novel crossmodal associations (Gonzalo *et al.*, 2000).

In extrapolating to BOLD effects from electrophysiological data, superadditive effects can be most easily interpreted in terms of local cerebrovascular reflections of the raised energy demands accompanying increased neuronal firing rate (Heeger *et al.*, 1999). To show genuine analogy between single cell work and fMRI data, it would be desirable if future experiments were carried out using stepwise increments in unimodal stimulus level/effectiveness to establish the true form of the relationship between the degree of crossmodal enhancement and unimodal stimulation. Interpretation of response depression is more problematic. In its electrophysiological context, it can indicate an active inhibition, which could lead to an increased BOLD effect if sampled at the synapses of the inhibitory cells, or decreased BOLD effect at the inhibited cells (Waldvogel *et al.*, 2000). This question is difficult to resolve at present due to the crude resolution of fMRI but we must infer that the cell populations sampled in (for example) the superior colliculus have reduced energy demand in the presence of unmatched stimuli.

In summary, we believe that the use of superadditive and suppressive effects to detect regions of intersensory integration may have considerable methodological advances in imaging structures where other approaches to the detection of such sites may encounter difficulties. For example, the superior colliculus is located in a region where respiratory or cardiac pulsation artefacts may render imaging difficult and where motion effects may reduce power to detect responses. Under these circumstances, the ability to detect activation in this structure by virtue of its response to any one sensory modality in isolation may be difficult. The large size of the positive and negative interactions observed in response to matched and mismatched bimodal stimuli may ease the detection problem considerably and overcome the requirement for a region of interest approach. Such methodology may also significantly enhance the power of functional neuroimaging to identify the brain networks involved in the process of multisensory integration.

ACKNOWLEDGMENTS

G.A.C. is supported by the Medical Research Council of Great Britain. P.C.H. acknowledges the MRC Interdisciplinary Research Centre for Cognitive Neuroscience, University of Oxford. Thanks to David Harwood and Andrew Matheson, Oxford Medical Illustrations, John Radcliffe Hospital, Oxford, and Chris Andrew at the Institute of Psychiatry, London, for technical support.

REFERENCES

- Andersen, R. A., Snyder, L. H., Bradley, D. C., and Xing, J. 1997. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annu. Rev. Neurosci.* **20**: 303–330.
- Bruce, C., Desimone, R., and Gross, C. G. 1981. Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J. Neurophysiol.* **46**: 369–384.
- Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Rabe-Hesketh, S., Janot, N., David, A. S., Mellers, J. D. C., Howard, R., and Sham, P. 1996. Statistical methods of estimation and inference for functional MR image analysis. *Magn. Res. Med.* **35**: 261–277.
- Bushara, K. O., Grafmann, J., and Hallett, M. 2001. Neural correlates of auditory-visual onset asynchrony detection. *J. Neurosci.* **21**(1): 300–304.
- Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D., and David, A. S. 1997. Activation of auditory cortex during silent lipreading. *Science* **276**: 593–596.
- Calvert, G. A., Brammer, M., Bullmore, E., Campbell, R., Iversen, S. D., and David, A. 1999. Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* **10**: 2619–2623.
- Calvert, G. A., Campbell, R., and Brammer, M. J. 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in human heteromodal cortex. *Curr. Biol.* **10**: 649–657.
- Colby, C. L., and Duhamel, J. R. 1991. Heterogeneity of extrastriate visual areas and multiple parietal areas in the macaque monkey. *Neuropsychologia* **29**: 517–537.
- Corbetta, M., Miezin, F. M., Shulman, G. L., and Petersen, S. E. 1991. Selective attention modulates extrastriate visual regions in humans during visual feature discrimination and recognition. *Ciba Found. Symp.* **163**: 165–175.
- Davis, K. D., Hutchison, W. D., Lozano, A. M., Tasker, R. R., and Dostrovsky, J. O. 2000. Human anterior cingulate cortex neurons modulated by attention-demanding tasks. *J. Neurophysiol.* **83**: 3575–3577.
- Driver, J. 1996. Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature* **2**(381): 66–68.
- Driver, J., and Spence, C. 1998. Attention and the crossmodal construction of space. *Trends Cogn. Sci.* **3**(3): 310–317.
- Eimer, M. 1999. Can attention be directed to opposite locations in different modalities? An ERP study. *Clin. Neurophysiol.* **110**: 1252–1259.
- Eimer, M., and Schroger, E. 1998. ERP effects of intermodal attention and cross-modal links in spatial attention. *Psychophysiology* **35**: 313–327.
- Foxe, J. J., Morocz, I. A., Murray, M. M., Higgins, B. A., Javitt, D. C., and Schroeder, C. E. 2000. Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. *Cogn. Brain Res.* **10**: 77–83.
- Frens, M. A., Van, O. A., and Van-der, W. R. 1995. Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Percept. Psychophys.* **57**: 802–816.
- Giard, M. H., and Peronnet, F. 1999. Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *J. Cogn. Neurosci.* **11**: 473–490.
- Gonzalo, D., Shallice, T., and Dolan, R. 2000. Time-dependent changes in learning audiovisual associations: A single-trial fMRI study. *Neuroimage* **11**: 243–255.
- Heeger, D. J., Boynton, G. M., Demb, J. B., Seidemann, E., and Newsome, W. T. 1999. Motion opponency in visual cortex. *J. Neurosci.* **19**(16): 7162–7174.
- Hikosaka, K., Iwai, E., Saito, H., and 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.
- Hughes, H. C., Reuter, L. P., Nozawa, G., and Fendrich, R. 1994. Visual-auditory interactions in sensorimotor processing: Saccades versus manual responses. *J. Exp. Psychol. Hum. Percept. Perform.* **20**: 131–153.
- Jay, M. F., and Sparks, D. L. 1984. Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature* **309**: 345–347.
- Jones, E. G., and Powell, T. P. 1970. An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain* **93**: 793–820.
- Kadunce, D. C., Vaughan, J. W., Wallace, M. T., Benedek, G., and Stein, B. E. 1997. Mechanisms of within- and cross-modality suppression in the superior colliculus. *J. Neurophysiol.* **78**: 2834–2847.
- King, A. J., and Palmer, A. R. 1985. Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus. *Exp. Brain Res.* **60**: 492–500.
- Lewis, J. W., Beauchamp, M. S., and DeYoe, E. A. 2000. A comparison of visual and auditory motion processing in human cerebral cortex. *Cereb. Cortex.* **10**(9): 888.
- Lewkowicz, D. J. 1996. Perception of auditory-visual temporal synchrony in human infants. *J. Exp. Psy. Human Perception and Performance.* **22**(5): 1094–1106.
- Macaluso, E., Frith, C., and Driver, J. 2000a. Selective spatial attention in vision and touch: Unimodal and multimodal mechanisms revealed by PET. *J. Neurophysiol.* **83**: 3062–3075.
- Macaluso, E., Frith, C., and Driver, J. 2000b. Modulation of human visual cortex by crossmodal spatial attention. *Science* **253**: 85–87.
- Meredith, M. A., and Stein, B. E. 1983. Interactions among converging sensory inputs in the superior colliculus. *Science* **221**: 389–391.
- Meredith, M. A., and Stein, B. E. 1986. Spatial determinants of multisensory integration in cat superior colliculus neurons. *J. Neurophysiol.* **75**: 1843–1857.
- Mesulam, M. M. 1998. From sensation to cognition. *Brain* **121**(Pt 6): 1013–52.
- Mesulam, M. M., and Mufson, E. J. 1982. Insula of the old world monkey. III: Efferent cortical output and comments on function. *J. Comp. Neurol.* **212**: 38–52.
- Miller, J. O. 1982. Divided attention: Evidence for coactivation with redundant signals. *Cogn. Psy.* **14**: 247–279.
- Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. 1990. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA* **87**: 9868–9872.
- Pandya, D. N., and Yeterian, E. H. 1985. Architecture and connections of cortical association areas. In: *Cerebral Cortex* (A. Peters and E. G. Jones, Eds.), pp. 3–61. Plenum Press, New York.
- Radeau, M. 1994. Auditory-visual spatial interaction and modularity. *Curr. Psychol. Cogn.* **13**: 3–51.
- Raij, T., Uutela, K., and Hari, R. 2000. Audiovisual integration of letters in the human brain. *Neuron* **28**(2): 617–25.

- Sekuler, R., Sekuler, A. B., and Lau, R. 1997. Sound alters visual motion perception. *Nature* **385**: 308.
- Seltzer, B., and Pandya, D. N. 1989. Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. *J. Comp. Neurol.* **281**: 97–113.
- Spence, C., and Driver, J. 1996. Audiovisual links in endogenous covert spatial attention. *J. Exp. Psychol. Hum. Percept. Perform.* **22**: 1005–1030.
- Stein, B. E. 1998. Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Exp. Brain Res.* **123**: 124–135.
- Stein, B. E., and Meredith, M. A. 1993. *Merging of the Senses*. MIT Press, Cambridge.
- Stein, B. E., Meredith, M. A., Huneycutt, W. S., and McDade, L. 1989. Behavioural indices of multisensory integration: Orientation to visual cues is affected by auditory stimuli. *J. Cogn. Neurosci.* **1**: 12–24.
- Talairach, J., and Tournoux, P. 1988. Co-planar stereotactic atlas of the human brain. Thieme, Stuttgart.
- Waldvogel, D., van Gelderen, P., Muellbacher, W., Ziemann, U., Immisch, I., and Hallett, M. 2000. The relative metabolic demand of inhibition and excitation. *Nature* **406**: 995–998.
- Wallace, M. T., Meredith, M. A., and Stein, B. E. 1992. Integration of multiple sensory modalities in cat cortex. *Exp. Brain Res.* **91**: 484–488.
- Wallace, M. T., Wilkinson, L. K., and Stein, B. E. 1996. Representation and integration of multiple sensory inputs in primate superior colliculus. *J. Neurophysiol.* **76**: 1246–1266.
- Wallace, M. T., and Stein, B. E. 2000. Onset of cross-modal synthesis in the neonatal superior colliculus is gated by the development of cortical influences. *J. Neurophysiol.* **83**(6): 3578–3582.
- Watanabe, J., and Iwai, E. 1991. Neuronal activity in visual, auditory and polysensory areas in the monkey temporal cortex during visual fixation task. *Brain Res. Bull.* **26**: 583–592.
- Watson, J. D., Myers, R., Frackowiak, R. S., Hajnal, J. V., Woods, R. P., Mazziotta, J. C., Shipp, S., and Zeki, S. 1993. Area V5 of the human brain: Evidence from a combined study using positron emission tomography and magnetic resonance imaging. *Cereb. Cortex.* **3**: 79–94.
- Welch, R. B., and Warren, D. H. 1986. Intersensory interactions. In *Handbook of Perception and Human Performance* (K. R. Kaufman and J. P. Thomas, Eds.), Vol. 1, pp. 1–36. Wiley.
- Wilkinson, L. K., Meredith, M. A., and Stein, B. E. 1996. The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. *Exp. Brain Res.* **112**: 1–10.