Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex

Gemma A. Calvert*, Ruth Campbell† and Michael J. Brammer‡

Background: Integrating information from the different senses markedly enhances the detection and identification of external stimuli. Compared with unimodal inputs, semantically and/or spatially congruent multisensory cues speed discrimination and improve reaction times. Discordant inputs have the opposite effect, reducing performance and slowing responses. These behavioural features of crossmodal processing appear to have parallels in the response properties of multisensory cells in the superior colliculi and cerebral cortex of non-human mammals. Although spatially concordant multisensory inputs can produce a dramatic, often multiplicative, increase in cellular activity, spatially disparate cues tend to induce a profound response depression.

Results: Using functional magnetic resonance imaging (fMRI), we investigated whether similar indices of crossmodal integration are detectable in human cerebral cortex, and for the synthesis of complex inputs relating to stimulus identity. Ten human subjects were exposed to varying epochs of semantically congruent and incongruent audio-visual speech and to each modality in isolation. Brain activations to matched and mismatched audio-visual inputs were contrasted with the combined response to both unimodal conditions. This strategy identified an area of heteromodal cortex in the left superior temporal sulcus that exhibited significant supra-additive response enhancement to matched audio-visual inputs and a corresponding sub-additive response to mismatched inputs.

Conclusions: The data provide fMRI evidence of crossmodal binding by convergence in the human heteromodal cortex. They further suggest that response enhancement and depression may be a general property of multisensory integration operating at different levels of the neuroaxis and irrespective of the purpose for which sensory inputs are combined.

Addresses: *Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK. †Department of Human Communication Science, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PG, UK. ‡Departments of Biostatistics and Computing, Institute of Psychiatry, London SE5 8AF, UK.

Correspondence: Gemma A. Calvert
E-mail: gemma@fmrib.ox.ac.uk

Received: **15 December 1999**
Revised: **8 February 2000**
Accepted: **24 March 2000**

Published: **16 May 2000**

Current Biology 2000, 10:649–657

0960-9822/00/$ – see front matter

Background

All higher organisms are equipped with several sensory systems, each tuned to a distinct form of energy and providing a unique window through which to experience the environment. Such multisensory capacity confers considerable behavioural flexibility as it permits the substitution of one sensory channel for another when necessary. For example, in darkness, auditory or tactile cues can supplement the impoverished visual input. Information from the different sensory streams can also be combined, providing information about the environment that is unavailable from any single modality, enhancing perception and reducing the ambiguity of external events.

Behavioural responses to semantically congruent and/or spatially coincident multisensory inputs in close temporal proximity exhibit lower thresholds and reduced reaction times than their unimodal counterparts [1–3]. Incongruent inputs have the opposite effect, slowing response times and producing perceptual anomalies [4–6]. This pattern of crossmodal enhancement and decrement in behaviour, examples of which abound in the human literature (for a review, see [7]), appears to be mirrored by the response properties of multisensory cells in non-human mammals [8]. These have been best characterised in the superior colliculus [9–11], a structure concerned with orientation and attentive behaviours, and more recently in the cerebral cortex [12,13]. Multisensory stimuli in spatial correspondence evoke a dramatic increase in firing rate in many of these cells [9,11,12] whereas spatially disparate cues produce response depression or no interaction [14,15]. These features of crossmodal processing in multisensory cells depend in part on the possession of overlapping sensory receptive fields [9,14]. It is important to note that not all multisensory cells that respond to stimulation in more than one modality necessarily integrate this information but, those that do, possess distinctive properties. Specifically, information obtained from more than one modality is transformed into an integrated product to produce an outcome that no longer resembles the unimodal inputs. Hence, the enhancement in cellular activity induced by spatially congruent multisensory cues is often

supra-additive (that is, greater than the sum of the individual inputs) and maximal in the presence of inputs that are minimally effective in isolation [9].

These parallels between the behavioural consequences of multisensory integration, evident in both man and other mammals, and the physiological indices of binding in multisensory cells in non-human mammals, raise some obvious questions. Firstly, does the principle of crossmodal binding by convergence onto multisensory neurons form a general physiological basis for intersensory synthesis that also extends to humans? Secondly, does this principle generalise to the binding of multisensory inputs relating to stimulus identity, as well as location, and presumably mediated in the cortex rather than the superior colliculus?

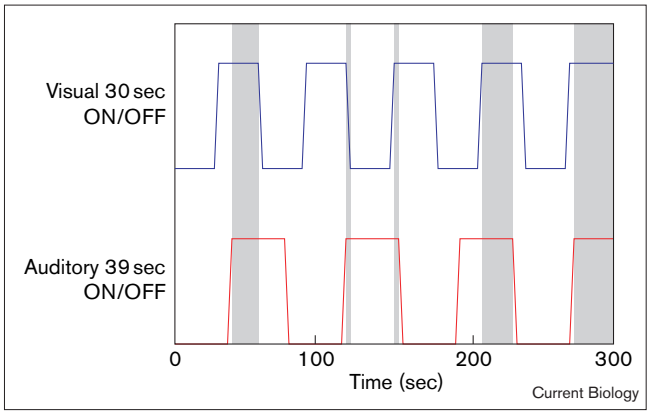
In non-human primates, areas of putative 'heteromodal' cortex have been identified on the basis that they receive inputs from multiple sensory modalities [16,17] and contain cells responsive to stimulation in more than one modality [18,19]. Nevertheless, ablation studies have failed to demonstrate a clear role for these areas in crossmodal performance (see [20] for a review). In humans, convergence of multiple sensory inputs in analogous regions is suggested on neuroanatomical and neuropsychological grounds [21] and from a recent study involving event-related potentials [22]. Other groups, however, have provided theoretical and empirical support for the view that there are no cortical convergence regions in which neuronal populations integrate information from the different sensory modalities or even from different submodalities [23,24]. One possible alternative is that crossmodal binding in man may be achieved instead by synchronised processing of the sensory inputs in their respective unimodal cortices [25], analogous to models of feature binding in the visual modality [26]. Such hypotheses, however, raise the question as to why afferents from the different modalities converge in heteromodal cortex, yet play no apparent part in integrating sensory information.

We conducted a functional magnetic resonance imaging (fMRI) study designed to detect evidence of multisensory integrative areas in the human cerebral cortex during the processing of semantically congruent and incongruent auditory and visual speech. These stimuli provide an excellent starting point for the investigation of crossmodal mechanisms in man for several reasons. Firstly, the behavioural enhancements and decrements produced by matched and mismatched audio-visual speech inputs have been extensively documented in the literature (for a review, see [27]). Secondly, there has been a plethora of research into the processing stage at which these speech signals combine. The established conclusion from these human speech perception studies is that the audible and visible evidences of speech converge at the phonetic level of speech processing [28,29]. Thus, a plausible psychological interpretation of our neurophysiological data can be attempted. Although fMRI does not, at present, allow the mapping of brain function at the level of individual multisensory cells, if these cells are present at a high enough density in a particular region, the local blood oxygen level dependent (BOLD) effect will allow us to examine their behaviour in response to a suitably chosen stimulus.

Results

Ten right-handed subjects were scanned in two experiments using a multiplexed fMRI paradigm (Figure 1) in which auditory (heard speech) and visual (mouthed speech) information was jointly presented in block format but at different epochs of alternation. Audible speech alternated with silence at 39 second intervals; visible speech alternated with a black screen at 30 second intervals, both throughout a total scanning time of 5 minutes. Subjectively, this design produced unpredictable periods of no stimulation, audible speech, visible speech, or simultaneous audio-visual speech with equal numbers of brain volumes collected in each of the four conditions. We predicted that this design would maximise the opportunity to detect BOLD signal changes paralleling the perceptual consequences experienced when switching from unimodal to bimodal speech (and vice versa) and minimise habituation and response attenuation by reducing expectancy. In experiment 1, the auditory and visual speech signals were congruent (lip movements dubbed precisely in time to the same audible story) and, in experiment 2, they were incongruent (different story being mouthed to that heard).

Figure 1



Experimental paradigm. The two crossmodal experiments employed a multiplexed paradigm in which the auditory and visual speech stimuli were presented at different frequencies of alternation, each against a 'no stimulation' rest condition (visual = 30 sec ON/OFF; auditory = 39 sec ON/OFF). This resulted in subjectively unpredictable periods of overlapping (audio-visual) and non-overlapping (auditory alone or visual alone) presentations of the two stimuli as well as periods of no stimulation (rest). Each of the resulting four conditions (auditory, visual, audio-visual, rest) were approximately equally distributed in this design. Areas of audio-visual costimulation are indicated by the grey bars.

By analogy with the response properties of multisensory integrative neurons in the superior colliculus, we hypothesised that brain areas involved in the audio-visual integration of speech should meet the following criteria: first, they should respond to both auditory and visual speech when separately presented (but see below); second, they should show BOLD responses that are significantly greater than the sum of the unimodal responses (supra-additive) during simultaneous presentation of congruent audio-visual inputs; and third, show significantly weaker (sub-additive) BOLD responses during presentation of incongruent inputs.

Brain areas meeting strict criteria for a site of multisensory integration

The only brain area that fulfilled all three strict criteria specified for a site of audio-visual speech integration was a cluster of eight voxels localised in the ventral bank of the superior temporal sulcus (STS) in the left hemisphere ($x = -49$, $y = -50$, $z = 9$; Figure 2). The bimodal response enhancement to congruent audio-visual speech in this cluster was 30–80% over that obtained by summing the auditory and visual responses. In contrast, incongruent audio-visual inputs reduced the response in these voxels to less than 50% of the summed unimodal responses. The size of the supra- and sub-additive effects in these eight voxels are illustrated in Figure 3 as a percentage change of the combined response to both unimodal presentations. These response characteristics of this region of the STS resembles those of multisensory integrative neurons in the superior colliculi and cerebral cortex of non-human mammals.

Although the application of such strict criteria provides reliable physiological evidence for the existence of multisensory integrative neurons in man, it may be overly conservative, masking the true spatial extent of the area/s involved in the crossmodal integration of auditory and visual speech. Electrophysiological studies in the cat superior colliculi have shown that, in some instances, supra-additive response enhancements to bimodal audio-visual stimuli can be elicited in multisensory cells even though their responses to the individual (auditory or visual) components may be weak or even below threshold for firing [8]. In a further stage of analysis, we thus relaxed our criteria by excluding the necessity for statistically significant unimodal responses to both auditory and visual inputs across both experiments while retaining the need to demonstrate both supra- and sub-additivity in response to semantically congruent and incongruent multisensory inputs, respectively. This manipulation did not result in any change in the location or size of the area already identified nor were there any additional areas showing significant supra- and sub-additivity under these criteria.

Localisation of supra-additive responses to matched audio-visual stimuli in experiment 1

Incongruent bimodal inputs are known to depress responses in multisensory colliculi neurons. These effects have, however, been observed less frequently than the corresponding response enhancement to congruent inputs and may be dependent on the presence of specific properties such as inhibitory surrounds [30]. By relaxing our criteria further to exclude the necessity to demonstrate sub-additivity, we were able to identify brain regions showing a supra-additive response enhancement to congruent audio-visual speech (Table 1; Figure 4, upper rows of each panel). These areas could be compared with those previously identified in our laboratory during the presentation solely of matched audio-visual speech stimuli and engaging a simple block design paradigm [31].

Significant supra-additive response enhancements to congruent audio-visual speech inputs were detected bilaterally along the posterior poles of the middle occipital gyri: Brodmann Area (B.A.) 18/19; extending anteriorly in the right hemisphere into the occipito-temporal junction (B.A. 19/37). This latter region corresponds to the visual motion area V5.

Figure 2

The cluster of voxels from the group averaged data in **(a)** axial and **(b)** coronal sections localised to the ventral bank of the left superior temporal sulcus ($x = -49$, $y = -50$, $z = 9$), which meet the criteria specified for a site of integration for auditory and visual speech signals. The images are displayed in radiological convention so that the left of the image corresponds to the right hemisphere.

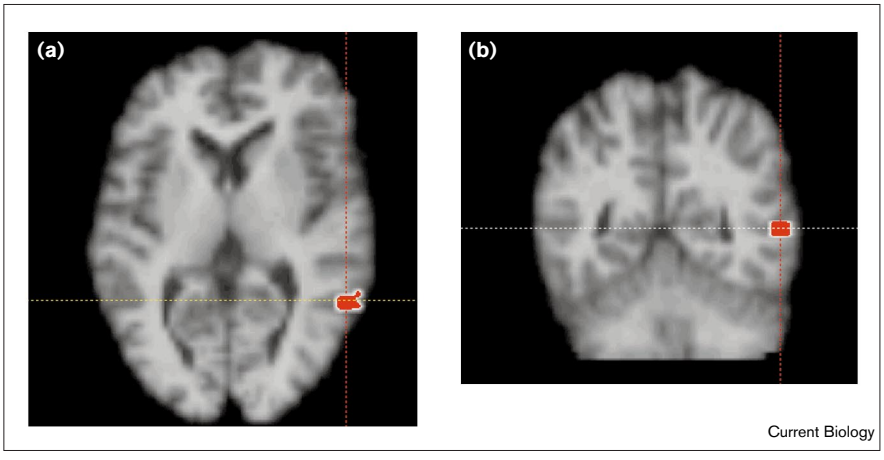
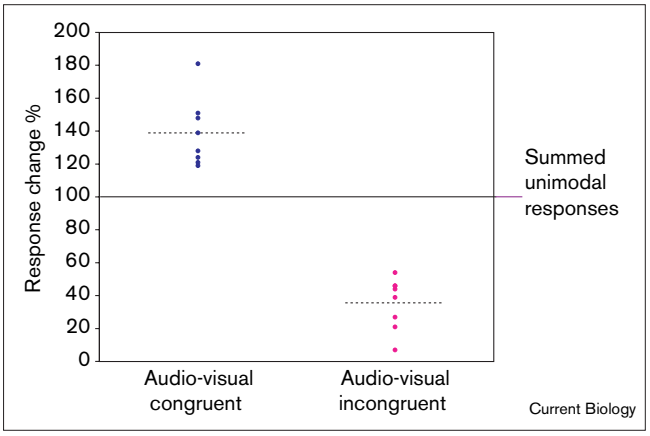


Figure 3



The responses to congruent and incongruent audio-visual speech presentation in each of the eight voxels in the STS cluster shown in Figure 2. The data shown are median responses over the ten subjects in the group. The levels of response to bimodal audio-visual speech were divided by the sum of the responses to unimodal auditory and visual stimulation and the results presented as a percentage (unimodal auditory + unimodal visual = 100). The mean levels of response to bimodal stimulation in the congruent and incongruent speech experiments are shown as dotted lines.

In the temporal lobe, analogous BOLD enhancements were also observed in the left and right superior temporal sulci (B.A. 22/21) and in left primary auditory cortex (B.A. 41/42) localised along Heschl's gyrus. Finally, supra-additive effects were also identified within the left middle frontal gyrus (B.A. 6/8) and right inferior parietal lobule (B.A. 40).

Localisation of sub-additive responses to mismatched audio-visual stimuli in experiment 2

The largest cluster of voxels showing significant sub-additive responses to incongruent auditory and visual stimulation was located in the left STS (B.A. 22/21), partially overlapping the strongest focus of supra-additivity. Other areas showing significant sub-additive responses (Table 2; Figure 4, lower rows of each panel) were detected bilaterally in the right and left inferior frontal region (B.A. 44/45), the premotor cortex (B.A. 6), the right superior temporal gyrus (B.A. 22) and the anterior cingulate gyrus (B.A. 32).

Discussion

By exploiting the BOLD response as an indicator of neuronal activity, we have identified a region in the posterior ventral bank of the left STS that displays response properties analogous to those of multisensory integrative cells. Specifically, voxels in this region exhibited a significant supra-additive response to semantically congruent audio-visual inputs, and corresponding sub-additive response to incongruent audio-visual inputs. Although multisensory cells have so far only been identified in non-human mammals, the observation that similar behaviour can be observed in human cerebral cortex under appropriate crossmodal conditions suggests that the principle of crossmodal binding by convergence onto multisensory neurons is also applicable in humans. The fact that it is possible to detect such indices of multisensory integration using fMRI suggests that there must be a substantial number of these integrative cells in this region of the STS to produce an analogous BOLD response.

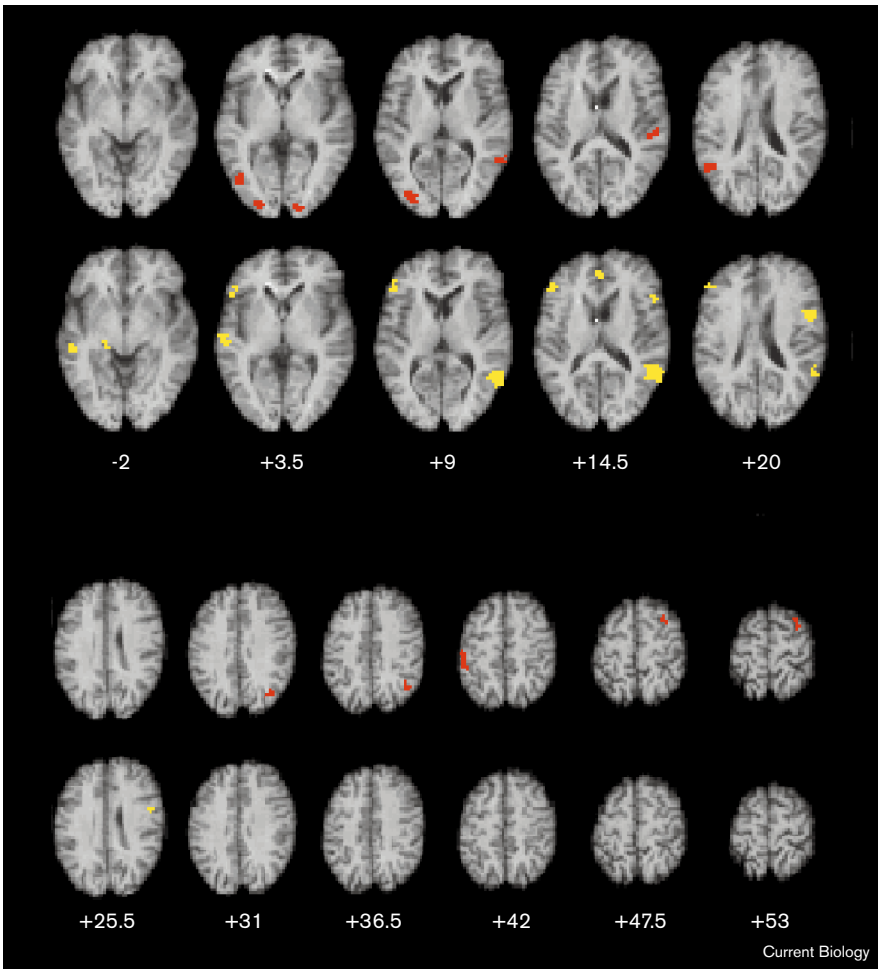
Data from non-human primates have highlighted the STS as a plausible site of multisensory integration. Neuroanatomical studies have identified areas within the STS that receive convergent inputs from visual, auditory and somatosensory cortices [16,17]. Electrophysiological data have shown that the cortex in this region contains cells responsive to stimulation in more than one sensory modality [18,19,32] and others sensitive to the sight of biologically meaningful actions, including head, eye and mouth movements [33,34]. Although homology between primate and human STS cannot automatically be assumed, evidence from human research also supports a role for the STS

Table 1

Brain areas showing supra-additive response enhancement to congruent audio-visual inputs.

Side	Brain region	B.A.	*x*	*y*	*z*	Cluster size	Effect size	*p* value
R	Fusiform gyrus	19	35	−70	−13	25	2.0	< 0.000001
R	Middle occipital gyrus	19	29	−84	7	23	2.3	< 0.000001
L	Middle frontal gyrus	6/8	−27	11	51	16	1.7	0.00009
R	Inferior parietal lobule	40	50	−32	42	16	1.9	< 0.000001
R	Occipito-temporal gyrus	37	43	−66	4	11	2.0	< 0.000001
R	Superior temporal sulcus	22/21	48	−55	20	10	1.9	< 0.000001
L	Heschl's gyrus	41/42	−48	−24	15	9	1.6	0.00009
L	Cuneus	17	−12	−92	4	8	1.6	0.00009
L	Superior temporal sulcus	22/21	−53	−48	9	8	1.5	0.00001

Figure 4



The locations of the supra-additive responses to congruent speech (red voxels, upper row of each panel) and sub-additive responses to incongruent speech (yellow voxels, lower row of each panel) in 11 axial sections centred on planes from 2 mm below the anterior–posterior commisural plane to 53 mm above it (indicated below each pair of images). The data are shown superimposed on a high-resolution anatomical image. Only activations with a voxel-wise type I error ($p \leq 0.0001$) are shown.

as a site of audio-visual integration. Studies using positron emission tomography (PET) and fMRI have shown that the STS is activated during both auditory [35–37] and visual [37,38] speech perception and, importantly, appears to be preferentially stimulated by phonetic features [39]. This is consistent with psychophysical data indicating that the audible and visible components of speech are integrated at the phonetic level, before lexical discrimination. Such conclusions derive from studies of normal audio-visual speech perception (for example, [28,29,40]), from studies of the visual bases of susceptibility to McGurk effects [41] (for example, lexical, semantic and syntactic processes have little effect on the strength of these illusions) as well as from studies confirming that infants perceive audio-visual speech in an integrated fashion (that is, they are susceptible to McGurk effects even though they have not yet learnt the meaning of words [42]).

Nevertheless, can we reconcile the integration of auditory and visual speech that occurs at an early stage of speech processing (that is, phonetic rather than lexical or semantic) with evidence of convergence in an area of heteromodal cortex (the STS) normally associated with the processing of signals that have undergone considerable elaboration [21]? The conventional explanation is that phonetic features are themselves highly elaborated signals in comparison to simple tones [28]. Integration of these modality-specific inputs in STS may thus be followed by an interactive process whereby the activated phonetic features are matched to specific lexical items (for example, see [43]) in upstream heteromodal areas such as the middle temporal gyrus (MTG), which abuts the superior temporal gyrus in the fundus of the STS. Such a model is consistent with imaging data indicating that the MTG is preferentially stimulated by lexical and semantic processing [39].

That laterality differences in crossmodal binding may be present in the STS is suggested by comparison of the present findings with previous neuroimaging data. Using fMRI, Puce and colleagues [44] reported bilateral activation (greater in the right hemisphere) of the STS when subjects passively viewed mouth movements. This centroid of

Table 2

Brain areas showing sub-additive responses to incongruent audio-visual inputs.

Side	Brain region	B.A.	*x*	*y*	*z*	Cluster size	Effect size	*p* value
L	Superior temporal sulcus	22/21	–49	–50	13	59	3.2	< 0.000001
R	Inferior frontal gyrus	45/46	48	30	12	28	3.0	< 0.000001
L	Inferior frontal gyrus	44/6	–47	4	22	21	3.2	< 0.000001
R	Superior temporal gyrus	22	56	–21	2	20	2.9	< 0.000001
M	Anterior cingulate gyrus	32	3	42	15	7	3.1	< 0.000001
L	Inferior frontal gyrus	45	–49	20	15	5	2.4	< 0.000001

maximum activation was located at precisely the same *z* (+12) and *y* (–50) Talairach coordinates as those reported in the current study, but was 3 mm more medial and in the opposite hemisphere. These laterality differences are likely to be due to the presence in our study of communicative mouth movements (seen speech) and may reflect functional differences in the roles of the left and right STS in the crossmodal integration of speech and non-speech inputs. This distinction gains support from neuropsychological data. Campbell and colleagues [45] reported that a lesion in the left occipito-temporal region (including posterior STS) produced impaired lipreading skills. Conversely, damage to the analogous area in the right hemisphere left lipreading intact.

In experiments involving shifts between different modalities or combinations of stimuli, the possibility exists that some changes in activation may reflect differences in attention. However, this is very unlikely to explain the signal changes reported in the STS in the current study. In a recent study of selective attention to utterances [46], subjects were instructed to attend either to auditory or visual syllables that were presented in simultaneous streams but temporally offset to prevent binding. A number of brain areas, including the STS, showed significant activation. However, whereas auditory and visual cortices exhibited attentional modulation, no such effects were seen in the STS.

Although the present study was limited to the investigation of integrative sites during the perception of auditory and visual speech, it nevertheless contributes important fMRI data to the growing body of evidence from other sources implicating human heteromodal cortex in multisensory synthesis [22,39,46]. The findings contrast, however, with previous studies in non-human primates [47,48] and PET data in humans [24] which have failed to produce convincing evidence of crossmodal processing in the heteromodal cortex. One explanation for the discrepancy between these findings and the results of the present study may relate to the choice of experimental paradigm. Lesion studies in primates have tended to focus on the crossmodal transfer or matching of sensory inputs, where information perceived from different modalities and relating to two distinct objects is matched along some shared dimension (for example, size, shape, intensity). Such tasks, however, fail to meet the criteria for binding, namely that the inputs are perceived as relating to a single object or common event. Although crossmodal matching is clearly a phenomenon that may recruit multisensory neurons, it may have little to do with the integration of inputs from different modalities nor require the involvement of heteromodal sites.

The detection of such significant supra- and sub-additive responses in the STS suggests that it plays a key role in the integration of audio-visual speech. We cannot, however, exclude the possibility that other, less prominent, integrative regions also participate in the synthesis of these signals but at a level currently undetectable by fMRI and with the imposition of criteria explicitly designed to minimise the potential for false positives. Indeed, in the current study, brain areas other than the STS displayed weaker supra- or sub-additive effects in response to congruent or incongruent stimuli but not to both. The response enhancements observed in primary auditory and visual motion cortices (Figure 4, upper rows of both panel) to matched audio-visual inputs replicate the findings from our previous fMRI study showing that bimodal (congruent) audio-visual speech evokes a greater response in these regions than either unimodal component [37]. In the absence of known direct connections between the auditory and visual cortices, we argued that these enhancements must reflect the downstream consequences of crossmodal integration in heteromodal zones which are then subsequently back-projected to modulate both sensory-specific areas. The results from the present study indicate that the relevant heteromodal region lies within the left STS. We further proposed that the amplifications in signal intensity in the auditory and visual cortices are the physiological correlates of the perceptual gains experienced during multisensory signal combination: specifically, the

subjective experience of an improvement in 'hearing' when the speaker can be both seen and heard [31], and enhanced 'visual' attraction towards the sound source when confronted by multiple speakers [49].

The interpretation of sub-additive effects in response to incongruent stimuli is necessarily more complex as multisensory depression has received considerably less attention at the neuronal level than multisensory enhancement and has not been studied well in fMRI. As Table 2 shows, the largest region showing significant sub-additive effects was the left STS, where there was considerable overlap with the region showing significant supra-additivity (Figure 4, slice 3, rows 1 and 2 of upper panel). It should be noted that the BOLD signal in this region arises from the cumulative averaged activation of neuronal firing from different neuronal types (unisensory neurons, bimodal and multisensory integrative). Hence, on average, in the incongruent condition we see a signal greater than either modality alone but significantly less than the null hypothesis of a linear summation of the two unimodal auditory and visual activations. The anterior cingulate has been implicated in attentional processes [50] and the detection of a sub-additive response in this area may reflect attempts to deflect attention to one or other modality when the two are mutually interfering. Sub-additive effects in the left and right inferior frontal gyri (in and adjacent to Broca's area and its putative right homologue) and in left superior temporal cortex may arise as a consequence of inhibition in these language processing areas [39] caused by the conflicting auditory and visual speech inputs. This is consistent with the familiar experience of viewing a dubbed foreign movie in which the incongruent lip and mouth movements interfere with the perception of an otherwise clear auditory speech signal.

Conclusions

By modelling situations in which the audible and visible components of speech are synthesised to enhance or degrade perception, and using fMRI to measure the resulting BOLD effects, we have identified an area within the left STS that displays response properties characteristic of multisensory integrative cells hitherto only demonstrated in non-human mammals. Further experimentation with non-speech sounds, phonetic, syllabic and semantic stimuli are necessary to discriminate the nature of the features being integrated in the STS but these data clearly support the hypothesis that crossmodal binding of sensory inputs in man can be achieved by convergence onto multisensory cells localised in heteromodal cortex. The study further illustrates the utility of fMRI in exploring the haemodynamic correlates of a number of behavioural elements characteristic of intersensory processes. Such methods may prove to be of considerable value in elucidating the mechanisms of multisensory interactions in the human central nervous system.

Materials and methods

Subjects

Ten right-handed native English-speaking subjects (mean age 30.1, range 22–45 years; 5 males and 5 females) participated in the study. All subjects were in good health with no past history of psychiatric or neurological diseases and gave informed consent to the protocol that has been approved by the local Research Ethics Committee. Subjects had normal or corrected-to-normal (with contact lenses) visual acuity.

Design

All fMRI scans were conducted on the same day and the two experiments reported here were randomly interleaved between two analogous non-speech experiments (data not reported here) to avoid order effects. The stimuli comprised extracts from George Orwell's '1984' read aloud at a normal rate by a female English speaker. During the recording, the camera's field of view was restricted to the lower half of the speaker's face to minimise the influence of gaze and facial identity processing during the experiments. Visual stimuli were recorded on videotape and projected onto a screen located at the base of the scanner bed through a Proxima 8300 LCD projector. The stimuli were viewed through a mirror angled above the subject's head in the scanner. Auditory stimuli were presented from the audio output of a video recorder through a pneumatic headset designed to minimise interference from scanner noise. The sound level of the speech was ~95 dB with scanner noise attenuated to 80 dB.

Subjects were instructed to follow and comprehend the story and to maintain fixation on the back-projection screen for the duration of each experiment. Psychophysical research has shown that the integration of heard and seen speech signals occur pre-attentively and is thus immune to attempts to shift attention to one or other modality [29]. To avoid interference from memory encoding or rehearsal strategies, we chose a passive perception task that would most closely resemble the electrophysiological stimulation experiments reported by Stein and Meredith [9] to detect multisensory responses in mammals. This design further allowed us to present a continuous stream of speech rather than intermittent words, which would minimise bimodal/unimodal shift enhancements.

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; 100 T2*-weighted images depicting BOLD contrast [51] were acquired over 5 min at each of 14 near-axial non-contiguous 7 mm thick planes parallel to the intercommisural (AC-PC) line: TE = 40 msec, TR = 3 sec, 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 stereotactic space [52] (TE = 80 msec, TI = 180 msec, TR = 16 sec, in-plane resolution 3 mm, number of signal averages = 8).

Data analysis

In contrast to electrophysiological studies, which can determine response characteristics of single cells, BOLD effects in fMRI are representative of the averaged responses of tens or hundreds of millions of neurons. Thus, in an fMRI study, the fact that a single voxel shows significant responses to auditory and visual stimulation and simple summation of the responses with bimodal input does not indicate convergence of these inputs on the same cells. It could simply be the case that some cells in that region respond only to unimodal auditory stimulation and a different group of cells to unimodal visual stimulation. To identify auditory and visual interactions at sites of convergence, we used a linear modelling approach. This involved determining the BOLD responses to unimodal auditory and visual inputs and the interaction effects between the two. A significant positive interaction effect (supra-additivity) indicates that the response to bimodal stimulation is greater than that obtained by summing the unimodal responses. Such an effect should be observed in an area of bimodal integration in response to congruent or matched stimuli. In contrast, sub-additivity or negative interaction effects should be seen with unmatched stimuli.

Prior to time-series analysis, the data for each subject were pre-processed to remove low-frequency signal changes and minimise movement-related artefacts. The responses to the multiplexed audio-visual stimuli were then analysed by least-squares fitting a linear model on a voxel-wise basis of the form

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

where Y is the image intensity at time point t and CV, CA, CAV are the convolutions of the epochs of auditory alone (CA) , visual alone (CV), and simultaneous auditory-visual (CAV) stimulation with a Poisson function modelling a haemodynamic delay of 6 sec [53]. This modelling technique represents each condition (*a*, *v* or *av*) in only one column of the design matrix at a given time point as failing to do so would result in non-independence of the parameter estimates. It also permits estimation of responses to even short epochs (3–6 sec) of stimulation as encountered in the multiplexed design used in this study. After fitting the model, the parameters *m*, *v*, *a* and *av* characterise the mean image intensity (*m*), and the magnitudes of the visual (*v*) and auditory (*a*) responses alone and the response to simultaneous auditory-visual stimulation (*av*). ε_t is the residual error at time *t*. The statistical significance of any of the model parameters (*m*, *v*, *a* or *av*) and of the interaction effects I, which are equal to (*av* – (*v* + *a*)), can be assessed by comparison with the null distributions computed by fitting the same model repeatedly at each voxel following random permutation of the time series (to destroy any pattern of experimentally determined effect) and removal of residual autocorrelation. Combination of the resulting parameter estimates over all voxels in the images yields the distribution of values of each model parameter under the null hypothesis that there is no experimentally determined response. The critical value of each parameter for testing at any desired *p* value can then be computed from this distribution [54].

The voxel-wise observed and randomised estimates of *v*, *a* and *I* and the sinusoidal regression data were then transformed into standard stereotactic space as described by Brammer and colleagues [55] for the construction of group activation maps. Median images of supra-additive effects (significant positive values of *I*) were then computed from the congruent speech data and equivalent images of sub-additive effects (significant negative values of *I*) from the discordant speech data. Significance testing at group level was carried out using the randomised data as previously described [55]. Finally, overlap images between these two maps were computed following thresholding of each map at a voxel-wise type I error probability (*p*) of 0.001.

Given the small size of the cluster surviving these analyses, it was important to demonstrate that it could not arise by chance conjunction of clusters of type I errors in the supra- and sub-additive interaction maps. To test this possibility, we have analysed a null data set obtained, with identical acquisition parameters to those of the current study, from six individuals in which no experimental paradigm was imposed. We then repeated the analysis procedure for the experiments described in this study to make a 'null' supra-additive effect map. The distribution of cluster sizes in this map was determined repeatedly by random sampling, replacement (bootstrapping) and reanalysis of the subject group one hundred times. To make a very conservative test, we also employed a much more lenient threshold ($p = 0.01$) than that used to generate the current supra- and sub-additive interaction maps ($p = 0.001$). Even at this much more lenient voxel-wise type I error threshold, the 95% confidence limit on cluster size was six voxels (that is, the chance of occurrence of a single cluster of size $\geq$ six voxels was ≤ 0.05). The 95% confidence limit on cluster size at a voxel-wise type I error probability of 0.001 was four voxels. The null hypothesis that a cluster of eight voxels might occur in any one interaction effect map by chance during the current image processing and analysis procedures can thus be rejected at $p < 0.05$ even with a lenient voxel wise type I error rate of 0.01. Given this prior probability in one map, the possibility that a cluster of this size or greater would occur in a second independent experiment and be present at exactly the same location amongst a volume of > 21000 voxels can be discounted. The alternative hypothesis that the observed conjunction of supra- and sub-additive effects reflects the fact that the same brain area is differentially modulated in the two experiments is thus accepted.

Acknowledgements

G.A.C. is supported by the Medical Research Council of Great Britain. We thank David Harwood and Andrew Matheson at Oxford Medical Illustrations for technical support. Thanks also to Susan Iversen, Heidi Johansen-Berg, Peter Hansen and Roger Tootell for their helpful comments on the manuscript.

References

1. Hershenson M: **Reaction time as a measure of intersensory facilitation.** *J Exp Psychol* 1962, **63:**289-293.
2. Morrell LK: **Temporal characteristics of sensory interaction in choice reaction times.** *J Exp Psychol* 1968, **77:**14-18.
3. Frens MA, Van Opstal: **Spatial and temporal factors determine audio-visual interactions in human saccadic eye movements.** *Percept Psychophys* 1995, **57:**802-816.
4. Stein BE, Meredith MA, Huneycutt WS, McDade L: **Behavioural indices of multisensory integration: orientation to visual cues is affected by auditory stimuli.** *J Cog Neurosci* 1989, **1:**12-24.
5. McGurk H, Macdonald J: **Hearing lips and seeing voices.** *Nature* 1976, **264:**746-748.
6. Sekuler R, Sekuler AB, Lau R: **Sounds alter visual motion perception.** *Nature* 1997, **385:**308.
7. Welch RB, Warren DH: **Intersensory interactions.** In *Handbook of Perception and Human Performance Volume 1: Sensory Processes and Perception.* Edited by Kaufman KR, Thomas JP. New York: Wiley; 1986:1-36.
8. Stein BE: **Neural mechanisms for synthesizing sensory information and producing adaptive behaviors.** *Exp Brain Res* 1998, **123:**124-135.
9. Stein BE, Meredith MA: *Merging of the Senses*. Cambridge: MIT Press; 1993.
10. King AJ, Palmer AR: **Integration of visual and auditory information in bimodal neurones in the guinea-pig superior colliculus.** *Exp Brain Res* 1985, **60:**492-500.
11. Meredith MA, Stein BE. **Interactions among converging sensory inputs in the superior colliculus.** *Science* 1983, **221:**389-391.
12. Wallace MT, Meredith MA, Stein BE: **Integration of multiple sensory modalities in cat cortex.** *Exp Brain Res* 1992, **91:**484-488.
13. Schroeder CE, Lindsley RW, Specht C, Marcovici A, Smiley JF, Javitt DC: **Somatosensory input to auditory association cortex in macaques: an anatomical basis for multisensory integration.** *J Neurophysiol* 2000, in press.
14. Meredith MA, Stein BE: **Spatial determinants of multisensory integration in cat superior colliculus neurons.** *J Neurophysiol* 1996, **75:**1843-1857.
15. Wallace MT, Wilkinson LK, Stein BE: **Representation and integration of multiple sensory inputs in primate superior colliculus.** *J Neurophysiol* 1996, **76:**1246-1266.
16. Jones EG, Powell TP: **An anatomical study of converging sensory pathways within the cerebral cortex of the monkey.** *Brain* 1970, **93:**793-820.
17. Seltzer B, Pandya DN: **Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey.** *Brain Res* 1978, **149:**1-24.
18. Desimone R, Gross CG: **Visual areas in the temporal cortex of the macaque.** *Brain Res* 1979, **178:**363-380.
19. Hikosaka K, Iwai E, Saito H, Tanaka K: **Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey.** *J Neurophysiol* 1988, **60:**1615-1637.
20. Ettlinger G, Wilson WA: **Cross-modal performance: behavioural processes, phylogenetic considerations and neural mechanisms.** *Behav Brain Res* 1990, **40:**169-192.
21. Mesulam MM: **From sensation to cognition.** *Brain* 1998, **121:**1013-1052.
22. Giard MH, Peronnet F: **Auditory-visual integration during multimodal object recognition in humans: a behavioural and electrophysiological study.** *J Cog Neurosci* **11**:473-490.
23. Singer W, Gray CM: **Visual feature integration and the temporal correlation hypothesis.** *Annu Rev Neurosci* 1995, **18:**555-586.
24. Hadjikhani N, Roland PE: **Cross-modal transfer of information between the tactile and the visual representations in the human brain: a positron emission tomographic study.** *J Neurosci* 1998, **18:**1072-1084.
25. Von Stein A, Rappelsberger P, Sarnthein J, Petsche H: **Synchronization between temporal and parietal cortex during multimodal object processing in man.** *Cereb Cortex* 1999, **9:**137-150.

26. Treisman A: **The binding problem.** *Curr Opin Neurobiol* 1996, **6:**171-178.
27. Campbell R, Dodd BJ, Burnham D: *Hearing by Eye (II): Advances in the Psychology of Speechreading and Auditory-visual Speech*. Hove: Psychology Press; 1998.
28. Summerfield Q: **Lipreading and audio-visual speech perception.** *Phil Trans Roy Soc London* 1992, **335:**71-78.
29. Massaro DW: *Perceiving Talking Faces: From Speech Perception To a Behavioral Principle*. Cambridge, Massachusetts: MIT Press; 1998.
30. Kadunce DC, Vaughan W, Wallace MT, Benedek G, Stein BE. **Mechanisms of within- and cross-modality suppression in the superior colliculus.** *J Neurophysiol* 1997, **78:**2834-2847.
31. Calvert GA, Brammer M, Bullmore E, Campbell R, Iversen SD, David A: **Response amplification in sensory-specific cortices during crossmodal binding.** *Neuroreport* 1999, **10:**2619-2623.
32. Bruce C, Desimone R, Gross CG: **Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque.** *J Neurophysiol* 1981, **46:**369-384.
33. Hasselmo ME, Rolls ET, Baylis GC, Nalwa V: **Object-centred encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey.** *Exp Brain Res* 1989, **75:**417-429.
34. Perrett DI, Hietanen JK, Oram MW, Benson PJ: **Organization and functions of cells responsive to faces in the temporal cortex.** *Philos Trans R Soc Lond B Biol Sci* 1992, **335:**23-30.
35. Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T: **Human brain language areas identified by functional magnetic resonance imaging.** *J Neurosci* 1997, **17:**353-362.
36. Demonet JF, Chollet F, Ramsay S, Cardebat D, Nespoulous JL, Wise R, *et al.*: **The anatomy of phonological and semantic processing in normal subjects.** *Brain* 1992, **115:**1753-1768.
37. Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SCR, Mcguire PK, *et al.*: **Activation of auditory cortex during silent lipreading.** *Science* 1997, **276:**593-596.
38. Blasi V, Paulesu E, Mantovani F, Menoncello L, Giovanni UD, Sensolo S, *et al.*: **Ventral prefrontal areas specialised for lip-reading: a PET activation study.** *Neuroimage* 1999, **9:**S1003.
39. Binder JR: **Functional MRI of the language system.** In *Functional MRI*. Edited by Moonen CT, Bandettini P. Berlin, Heidelberg: Springer-Verlag; 1999:407-419.
40. Green KP: **The use of auditory and visual information during phonetic processing; implications for theories of speech perception.** In *Hearing By Eye II*. Edited by Campbell R, Dodd BM, Burnham D. Hove: Psychology Press; 1998:3-26.
41. Rosenblum LD, Saldaña HM: **Visual primitives for audiovisual speech integration.** *J Exp Psychol Hum Percept Perform* 1996, **22:**318-331.
42. Rosenblum LD, Schmuckler MA, Johnson JA: **The McGurk effect in infants.** *Percept Psychophys* 1997, **59:**347-357.
43. Elman JL, McClelland JL: **Speech perception as a cognitive process: the interactive activation model.** In *Speech and Language: Advances in Basic Research and Practice, Vol 102*. Edited by Lass NJ. New York: Academic Press; 1984:337-374.
44. Puce A, Allison T, Bentin S, Gore JC, McCarthy G: **Temporal cortex activation in humans viewing eye and mouth movements.** *J Neurosci* 1998, **18:**2188-2199.
45. Campbell R, Landis T, Regard M: **Face recognition and lipreading. A neurological dissociation.** *Brain* 1986, **109:**509-521.
46. Kawashima R, Imaizumi S, Mori K, Okada K, Goto R, Kiritani S, *et al.*: **Selective visual and auditory attention toward utterances — a PET study.** *Neuroimage* 1999, **10:**209-215.
47. Ettlinger G, Garcha HS: **Cross-modal recognition by the monkey: the effects of cortical removals.** *Neuropsychologia* 1980, **18:**685-692.
48. Horster W, Rivers A, Schuster B, Ettlinger G, Skreczek W, Hesse W: **The neural structures involved in cross-modal recognition and tactile discrimination performance: an investigation using 2-DG.** *Behav Brain Res* 1989, **33:**209-227.
49. Driver J: **Enhancement of selective listening of illusory mislocation of speech sounds due to lip-reading.** *Nature* 1996, **381:**66-68.
50. Posner MI, Petersen SE: **The attention system of the human brain.** *Annu Rev Neurosci* 1990, **13:**25-42.
51. Ogawa S, Lee TM, Kay AR, Tank DW: **Brain magnetic resonance imaging with contrast dependent on blood oxygenation.** *Proc Natl Acad Sci USA* 1990, **87:**9868-9872.
52. Talairach J, Tournoux P: *Co-planar Stereotactic Atlas of the Human Brain*. Stuttgart: Thieme; 1988.
53. Friston KJ, Holmes AP, Poline JB, Grasby PJ, Williams SC, Frackowiak RS, Turner R: **Analysis of fMRI time-series revisited.** *Neuroimage* 1995, **2:**45-53.
54. Bullmore E, Brammer M, Williams SC, Rabe HS, Janot N, David A, *et al.*: **Statistical methods of estimation and inference for functional MR image analysis.** *Magn Reson Med* 1996, **35:**261-277.
55. Brammer MJ, Bullmore ET, Simmons A, Williams SC, Grasby PM, Howard RJ, *et al.*: **Generic brain activation mapping in functional magnetic resonance imaging: a nonparametric approach.** *Magn Reson Imaging* 1997, **15:**763-770.

Because *Current Biology* operates a 'Continuous Publication System' for Research Papers, this paper has been published on the internet before being printed. The paper can be accessed from http://biomednet.com/cbiology/cub – for further information, see the explanation on the contents page.