

Marta Olivetti Belardinelli · Carlo Sestieri
Rosalia Di Matteo · Franco Delogu
Cosimo Del Gratta · Antonio Ferretti
Massimo Caulo · Armando Tartaro · Gian Luca Romani

Audio-visual crossmodal interactions in environmental perception: an fMRI investigation

Received: 20 May 2004 / Revised: 11 June 2004 / Accepted: 21 June 2004 / Published online: 14 July 2004
© Marta Olivetti Belardinelli and Springer-Verlag 2004

Introduction

In everyday life, the perception of surrounding events rarely takes place through a single sensory modality. Rather, perception is the result of the processing of information converging from the different senses. Behavioural research often proposes that the binding of different kinds of sensory input creates advantages in the detection, localization, and recognition of external events (King and Calvert 2001). However, the modalities with which multiple sensory cues deriving from the same object merge to form a coherent percept still represent a controversial topic in literature.

Research in this field, from the animal studies of Stein and Meredith (1993) to the latest human neuroimaging investigations (for a review see Calvert 2001), has shown the importance of the temporal and spatial congruence

of incoming stimuli in establishing crossmodal associations. Nevertheless, other characteristics, like semantic congruence, play a significant role in binding crossmodal associations, especially during the integration of information about complex objects. Thus the analysis of the semantic relationships established in crossmodal stimulation may be a promising instrument for exploring the neural substrates of multi-sensory integration.

At least two main theoretical views on the neural pathways involved in crossmodal processing have been proposed: the first one stresses the importance of multi-sensory cortical areas that receive projections from the different senses (Calvert 2001); the second view emphasizes the importance of the combined activity of the modality-specific cortices by means of synchronized firing (Ettlinger and Wilson 1990).

At the same time, an effort has been made to identify regions of the human brain, called heteromodal cortices, that receive afferents from different senses and are analogous to those described in animals. At the cortical level, these regions have been found in the superior temporal sulcus, in the intraparietal sulcus, and in the prefrontal and limbic cortices (Mesulam 1998; Calvert 2001). Heteromodal areas have also been found in sub-cortical structures, such as the superior colliculus (Stein and Meredith 1993). Recent neuroimaging investigations attempted to establish a relationship between different crossmodal tasks and the activation of specific heteromodal areas (Calvert 2001).

Experimental evidence reported the activation of the lateral temporal cortex in response to the integration of audio-visual information during recognition or identification tasks, as described by Calvert et al. (2000). These authors investigated the regions that exhibited supra-additive response enhancement to congruent audio-visual speech using functional magnetic resonance imaging (fMRI). They observed strong interaction effects in the left superior temporal sulcus.

Recently, Beauchamp et al. (2004) investigated the contribution of the superior temporal areas in the integration of visual and auditory information about

Edited by: Marie-Hélène Giard and Mark Wallace

M. Olivetti Belardinelli · F. Delogu
Department of Psychology,
University of Rome "La Sapienza",
P. Le Aldo-Moro, Rome, Italy

R. Di Matteo · M. Olivetti Belardinelli (✉) · F. Delogu
ECONA (Interuniversity Center for the Research on Cognitive
Processing in Natural and Artificial Systems),
University of Rome "La Sapienza",
P. Le Aldo-Moro, Rome, Italy
E-mail: marta.olivetti@uniroma1.it

C. Sestieri · C. Del Gratta · A. Ferretti · M. Caulo
A. Tartaro · G. L. Romani · R. Di Matteo
Department of Clinical Sciences and Biomedical Imaging,
University of Chieti-Pescara, Italy

C. Del Gratta · A. Ferretti · M. Caulo
A. Tartaro · G. L. Romani
Institute of Advanced Biomedical Technologies,
University of Chieti-Pescara, Italy

C. Del Gratta · A. Ferretti · G. L. Romani
INFN (National Institute for the Physics of Matter,
Coordinated Group of Chieti), L'Aquila, Italy

complex objects. They found interaction effects in audio-visual presentations in the posterior part of the superior temporal sulcus (pSTS) and in the middle temporal gyrus (MTG). However, when they directly compared congruent and non-congruent crossmodal stimuli, they observed that the effect found in pSTS/MTG was relatively weak. The authors concluded that the pSTS/MTG regions did not seem to be primarily sensitive to semantic congruency.

Nevertheless, other heteromodal regions have been associated with the binding of crossmodal congruent information. For example, Giard and Peronnet (1999), using ERPs, investigated the multisensory integration of meaningless non-verbal stimuli (simple sounds and images). They found interaction components in specific sensory cortices, followed by a component localized in the right fronto-temporal area, that seem to reflect multi-sensory processing. Gottfried and Dolan (2003) studied the relationships between vision and smell during an olfactory detection task. The authors found crossmodal effects in the anterior hippocampus and in the rostromedial orbitofrontal cortices, by varying the semantic congruency between odour-picture pairs. Since the hippocampus receives convergent inputs from different senses and has a role in associative memory, these results were interpreted as evidence that the hippocampus is involved in the reactivation of crossmodal semantic associations.

The influence of concurrent crossmodal cues on the recognition of auditory and visual targets was recently investigated in a behavioural study on environmental sounds and pictures of everyday experience (Delogu et al., submitted). In this study, the degree of semantic correspondence was manipulated so that cues could match or mismatch the meaning of the target. Velocity and accuracy of recognition of the visual and auditory targets were measured by using both synchronous (simultaneous presentations) and an asynchronous (priming) paradigm. Results showed an influence of crossmodal cues, in terms of facilitation effects, only for the visual targets on auditory ones but not vice versa.

The present study was designed in order to investigate, by means of fMRI, the neural correlates involved in this kind of crossmodal recognition. It was carried out by selecting 50 environmental stimuli from the set used in the behavioural study mentioned above (Delogu et al., submitted). Participants were exposed to two unimodal and two synchronous crossmodal conditions and instructed to recognize the stimuli in silence. We chose a passive paradigm to avoid the activation of cortical areas involved with motor and decision-making processes, since this type of paradigm has already been used in crossmodal studies (Calvert et al. 2000). The main difference between our study and the behavioural paradigm was that participants were not asked to attend to one modality, but to the event as a whole.

In particular, our aims were: (1) to detect cortical sites in which crossmodal presentations elicit a greater response than single modality presentations; and (2) to

provide evidence for differences in cortical processing of semantically congruent and non-congruent crossmodal information. These investigations can shed light on the way in which the brain links several pieces of information into a whole coherent event.

Methods

Participants

Thirteen right-handed male volunteers (mean age 22.8) participated in the study. All were in good health, free of neurological diseases and had normal hearing and vision. The study had been approved by the local ethics committee. Participants gave informed consent and were paid 25.5 Euro for participating.

Stimuli

Twenty-five visual stimuli and 25 auditory stimuli were used in the fMRI session. Twenty-four stimuli from the fMRI session and 24 (12 visual and 12 auditory) additional stimuli were used in a final recognition session.

Stimuli belonged to three categories: animals (i.e. cat, lion, horse), tools (i.e. airplane, camera, zip) and human acts (baby's cries, yawns, screams), and were equally distributed across the experimental conditions.

The visual stimuli consisted of coloured pictures, projected onto a screen located behind the scanner bed and viewed through a mirror above the subjects' heads. Pictures were presented at the centre of the screen at a visual angle of 14°.

Auditory stimuli consisted of environmental sounds, corresponding to the images, which were presented by means of a pneumatic headset, designed to minimize interference from scanner noise. Sounds were presented in stereo sound, at a frequency of 22,050 Hz, with a 16-bit resolution and variable amplitude of approximately 90 dB. Sound-length was 2.5 s and coincided with the image presentation time.

Procedure

The experiment consisted of three sessions: a preliminary practice session, an fMRI session and a final recognition session. The practice session was designed to familiarize participants with this kind of stimuli presentation. Participants performed trial versions of the task using a PC. They were informed about the experimental procedures, with the exception of the final recognition test, to avoid the development of memory strategies during the fMRI session. The stimuli presented during the practice session were not used in the other sessions.

The fMRI session lasted about 35 min, during which both functional and anatomical data were acquired. Participants were instructed to recognize the presented stimuli in silence, avoiding motor responses. During rest periods they were asked to keep gazing at the fixation point.

The final recognition test was designed to ascertain whether participants had attended to the stimuli during the fMRI session. Twenty-four stimuli used in the fMRI session and 24 new stimuli were presented using a PC. Participants were instructed to distinguish previously perceived stimuli from those that had not been presented by pressing two buttons on a keyboard.

fMRI experimental design

The experimental paradigm was created using E-Prime, Psychology Software Tools. It included four experimental conditions: audio, visual, corresponding audio-visual (Match), non-corresponding audio-visual (Mismatch). Each of the four conditions was repeated 25 times, using different randomized stimuli. A rest period followed, during which a grey screen with a black fixation point at the centre was presented for 14.5 s. Therefore, the paradigm was composed of 100 events (stimulus + rest) of 17 s each, forming an entire session of approximately 28 min. The order of events was random and differed for each participant (see Fig. 1).

fMRI image acquisition

Functional magnetic resonance imaging was performed with a Siemens Magnetom Vision scanner at 1.5 T, by means of T2*-weighted echo planar imaging (EPI) free induction decay (FID) sequences with the following parameters: TR 2 s, TE 60 ms, matrix size 64×64, FOV 256 mm, in-plane voxel size 4×4 mm, a flip angle 90°, slice thickness 6 mm and no gap. A total of 830 functional

volumes were acquired, consisting of 16 trans-axial slices, including the cortical regions of interest.

A high resolution structural volume was acquired at the end of the session via a 3D MPRAGE sequence with the following features: axial, matrix 256×256, FoV 256 mm, slice thickness 1 mm, no gap, in-plane voxel size 1×1 mm, flip angle 12°, TR 9.7 ms, TE 4 ms.

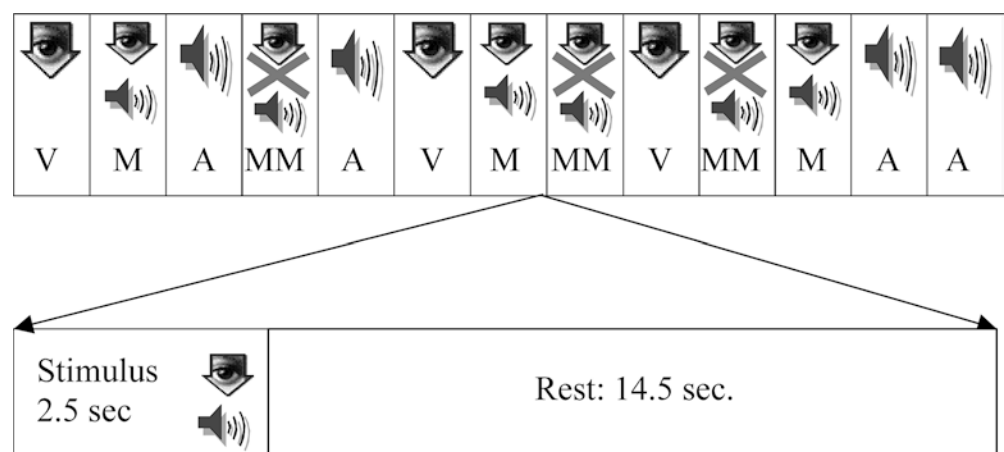
Data analysis

Data pre-processing and the statistical analysis of fMRI data were performed by means of the Brain Voyager 4.9 software (Brain Innovation, The Netherlands). Pre-processing included motion and slice scan time corrections, and the removal of linear trends from the time series. Data from three subjects were excluded from further data analysis because their movement had been too great. As a result, the analysis was conducted on a group of ten subjects. Functional 2D images were registered with the 3D high-resolution structural images and normalized in 3D Talairach Space (Talairach and Tournoux 1988). Functional volumes were resampled at a voxel size of 3×3×3 mm.

Statistical analysis was performed using a general linear model (Friston et al. 1995). The fMRI data was treated as an event-related design and regression coefficients were estimated for the four experimental conditions (match, mismatch, auditory and visual). Regressors were specified using a box-car wave form, convolved with an empirically derived hemodynamic response function (Boynton 1996).

Data analysis was performed in two steps. First, we carried out the group analysis for an initial selection of possible areas involved in crossmodal integration. Here, a conjunction analysis (Price and Friston 1997; Hadjikhani and Roland 1998; Calvert et al. 1999) was used in order to identify areas in which bimodal stimulation gave rise to greater responses than for the single modalities. In this type of statistical comparison the

Fig. 1 The diagram shows the time course of the fMRI experimental paradigm. *A* audio condition; *V* visual condition; *M* match condition (congruent audio and visual stimuli); *MM* mismatch condition (incongruent audio and visual stimuli)



computation can be described as $(AV-V) \cap (AV-A)$, where AV is the audio-visual, V is the visual and A is the auditory stimulation. The conjunction method, indicated by the \cap symbol, considers the common activation shared by two previously computed contrasts: the former, being between the crossmodal and the visual unimodal; the latter, between the crossmodal and the auditory unimodal condition. Each contrast was thresholded at a voxel level at $p < 0.000064$. A cluster size of at least four voxels was required. These thresholds and an estimate of the spatial correlation of voxels [Forman et al. (1995); 3dFWHM routine of AFNI package, Cox (1996)] were used for a Monte-Carlo simulation [AlphaSim routine of AFNI package, Cox (1996); Forman et al. (1995)] in order to assess the overall significance level (the probability of a false detection for the entire functional volume). In this way, we obtained a significance level of $p < 0.05$, corrected for multiple comparisons.

Second, we performed the individual subject analysis. The regions of interest (ROIs) were determined by considering the activated voxels during the match and mismatch conditions. The analysis of the BOLD response was then performed solely for the brain areas highlighted by the group analysis.

The mean time course of the fMRI signal of voxels belonging to a given ROI was investigated in order to evaluate the fMRI signal intensity variation during the four different conditions. The BOLD response was expressed as the signal's relative change (percentage of change) with respect to the baseline. It was calculated using the regression coefficients derived from the general linear model. Activation differences across conditions for these ROIs were assessed using analysis of variance (ANOVA) in which the Condition was the independent variable, and the BOLD response (percent of signal change) was entered as the dependent variable.

This analysis was conducted separately for matching and mismatching crossmodal stimuli. It should permit to identify brain areas specifically involved in the processing of semantically congruent and incongruent audio-visual information.

Results

Crossmodal effects in the match condition

The results of the conjunction analysis indicated that the main areas in which BOLD signals significantly increased during the crossmodal match condition, compared to the unimodal conditions, were the left parahippocampal gyrus, the left hippocampus of the medial temporal lobes and the lingual gyrus of the occipital cortex of both hemispheres. A crossmodal effect was also found among subcortical structures, in the thalamus, and bilaterally in the cerebellum (see Table 1).

We performed an ROI analysis on three areas: the left parahippocampal gyrus-hippocampus, the left and right lingual gyrus (see Fig. 2).

For the left parahippocampal gyrus, the ANOVA indicated a significant effect of the stimulus condition ($F_{3,27} = 7.340$; $p < 0.001$). The post-hoc analysis revealed that the match condition differed significantly from other conditions (LSD test $p < 0.01$), while the visual, auditory and mismatch conditions did not differ significantly.

The results of the ANOVA performed on the left lingual gyrus showed a significant effect of the stimulus condition ($F_{3,27} = 12.757$; $p < 0.0001$). The post-hoc analysis revealed that the match condition differed significantly from the mismatch, visual and auditory conditions (LSD test, $p > 0.05$).

Finally, the ANOVA performed on the right lingual gyrus showed a significant effect of the stimulus condition ($F_{3,27} = 12.173$; $p < 0.0001$), while the post-hoc analysis indicated that the match condition significantly differed from the mismatch, visual and auditory conditions (LSD test, $p < 0.05$).

Crossmodal effects in the mismatch condition

The most significant effect of the mismatch condition was observed in a region of the prefrontal cortex along

Table 1 Brain areas showing an effect of semantically congruent crossmodal pairs ($p < 0.000064$)

Side	Brain region	Talairach coordinates				Cluster size
		BA	x	y	z	
Left	Parahippocampal gyrus/hippocampus	28	-22	-21	-7	733
	Thalamus	28	16	-24	-6	414
Left	Lingual gyrus	18	-6	-74	-11	318
Right	Lingual gyrus	18	3	-77	-12	305
	Thalamus		1	-17	12	233
Right	Cerebellum		23	-55	-18	221
Left	Cerebellum		-13	-75	-20	126
Left	Lingual gyrus	18	-1	-70	4	87
Left	Cerebellum		-26	-56	-16	79
Right	Cerebellum		23	-77	-21	68
Right	Cerebellum		30	-65	-21	34
Left	Inferior occipital gyrus	18	-16	-89	-17	28

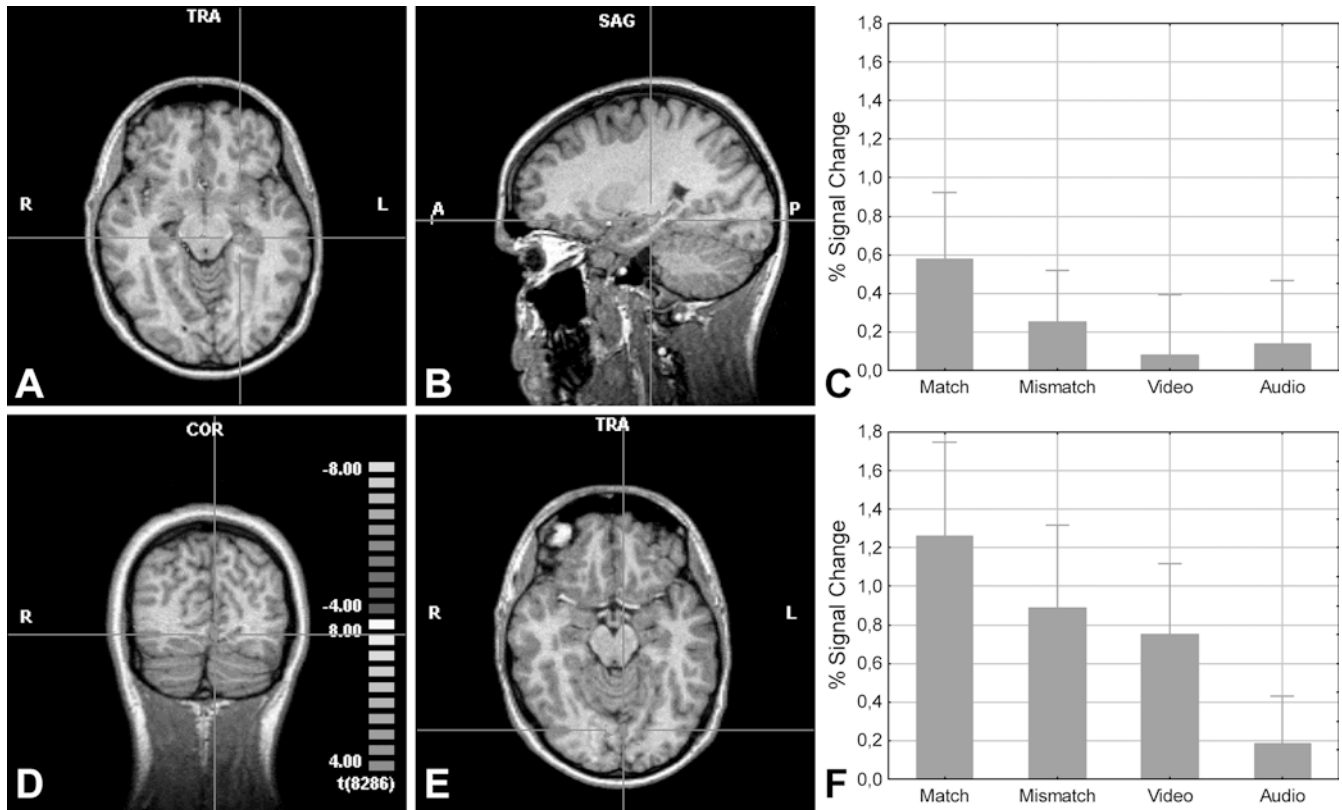


Fig. 2 **A** Transversal and **B** sagittal views of the activation in the left parahippocampal gyrus and hippocampus ($x=-22$, $y=-21$, $z=-7$) resulting from the conjunction analysis (match—audio) \cap (match—video) performed on the group. **C** Mean increase of BOLD signal during the match condition with respect to the other conditions (*vertical bars* indicate confidential intervals at 0.95). **D** Coronal and **E** transversal views of the activation in the left lingual gyrus ($x=-6$, $y=-74$, $z=-11$) resulting from the group analysis. **F** Mean increase of BOLD signal during the match condition with respect to the other conditions (*vertical bars* indicate confidential intervals at 0.95)

the left inferior frontal sulcus. Other minor effects were found bilaterally in the insula, and in the right inferior frontal sulcus (see Table 2).

A region of interest analysis was performed on the left prefrontal activation (see Fig. 3). The ANOVA showed a main effect of the stimulus condition ($F_{3,27}=6.772$; $p<0.005$) and the post-hoc analysis revealed that the percentage of signal change for the mismatch condition was significantly different from all the other conditions ($p<0.05$).

Discussion

The results of the comparison between the match and the two unimodal conditions allowed us to localize the areas that showed an effect for the combination of the two modalities and which may be a reasonable candidate for crossmodal associations. Activation was observed in two main areas: in the left medial temporal cortex, comprehending part of the parahippocampal gyrus and the hippocampus; and in the unimodal visual cortex of the occipital lobes. Anatomical and physiological evidence demonstrated that the hippocampus receives projections from several senses (Mesulam 1998), through converging sensory input from the entorhinal, perirhinal and parahippocampal cortices. A number of studies (Eichenbaum 1992; Bunsey and Eichenbaum 1996; Henke et al. 1997; Rombouts et al. 1997) have underlined the importance of the hippocampus in associative learning, especially in tasks where participants are required to link multiple stimuli. According to

Table 2 Brain areas showing an effect of semantically incongruent crossmodal pairs ($p<0.000064$)

Side	Brain region	BA	Talairach coordinates			Cluster size
			x	y	z	
Left	Inferior frontal sulcus	44/9	-43	13	30	826
Right	Insula		39	-11	0	33
Right	Inferior frontal sulcus	44/9	34	12	27	20
Right	Lingual gyrus	18	6	-80	-15	13
Left	Insula		-37	-22	7	10

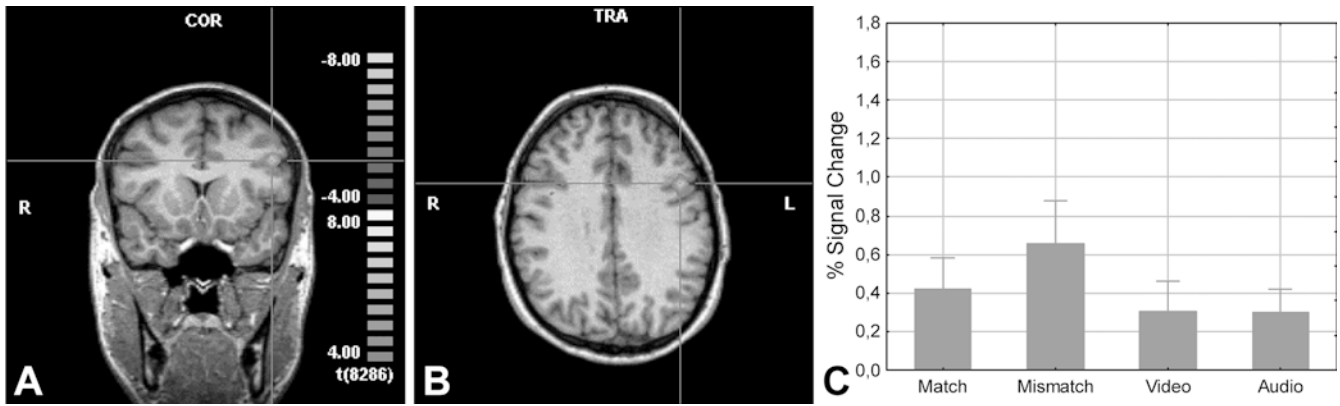


Fig. 3 **A** Coronal and **B** transversal views of the activation in the inferior frontal sulcus ($x = -43$, $y = 13$, $z = 30$) resulting from the conjunction analysis (mismatch—audio) \cap (mismatch—video) performed on the group. **C** Mean increase of BOLD signal during the mismatch condition with respect to the other (vertical bars indicate confidential intervals at 0.95)

Jackson and Schacter (2004), the hippocampal and medial temporal regions are specifically involved in the creation of durable links between individual items of information. Thus, the hippocampus seems to be particularly suited to registering and retrieving crossmodal associations. Interestingly, a model of explicit memory proposed by various authors (Mesulam 1990; McClelland 1995; Squire and Zola 1996; Nadel and Moscovitch 1997) suggests that facts and events are initially registered in multiple cortical sites. These reflect attribute- and category-specific aspects of the incoming information. Subsequently, the information is transmitted to the heteromodal cortices of the limbic system. According to the model, these heteromodal areas seem to play a critical role in developing links between different fragments of information, thus forming coherent multi-sensory experiences.

Further evidence indicates that the hippocampus is not just involved in overtly associative tasks but, more broadly, in the recollection of recently occurring events. For example, Stark and Squire (2000) found activity in the hippocampal region associated with recollection success during a crossmodal associative picture—name test. Gottfried and Dolan (2003) observed that the anterior hippocampus was only activated for semantically matching crossmodal pairs, even in the less demanding olfactory—visual detection task. These authors concluded that the obligatory reactivation of this kind of association does not require the engagement of intentional memory.

Our results suggest that the parahippocampal and hippocampal activation, observed mainly during the match condition, could reflect the involvement of these structures in recognition tasks, in which semantically congruent pieces of information about the same event are simultaneously presented. In this case, we can assume that the BOLD signal increase is related to the reactivation of crossmodal associations stored in

long-term memory, instead of to the recollection of recently occurring events.

However, the results are consistent with the view that the hippocampus and the adjacent cortex are generally involved during crossmodal recognition tasks that use meaningful stimuli. The effects detected in the occipital cortex were less striking than those previously considered, since the post-hoc analysis showed that the percentage of signal change for match and mismatch conditions was only marginally different ($p < 0.05$). Crossmodal effects in unimodal cortices have been observed in many studies on multisensory integration, using different paradigms and techniques (Calvert et al. 1999, 2000; Giard and Peronnet 1999; Molholm et al. 2002; Macaluso et al. 2000). These effects have been variably observed on their own or in combination with the activation of different structures, for example the prefrontal cortex (Giard and Peronnet 1999) or the superior temporal sulcus (Calvert et al. 2000). Recent experimental evidence has indicated the contribution of unimodal cortices during crossmodal tasks, in terms of feedback or back-projection mechanisms, originating from higher order cortices (Driver and Spence 2000; Macaluso et al. 2000). In our case, the activation observed in the medial temporal cortices, as a consequence of semantic matching, suggests that these areas may be involved in sending feedback signals to unimodal cortices. Nevertheless, our results do not allow us to firmly reject the hypothesis that the activity in unimodal areas is merely the result of the combination of two stimuli, simultaneously presented in the visual and the auditory modality, rather than the effect of semantic matching. In addition, it should be noted that no effect was found in the unimodal auditory cortices. This suggests that if an effect of semantic matching does exist in the unimodal cortices, it is confined to the visual cortex.

Since we were interested in separating the effects due to a mere temporal correspondence of incoming stimuli, from the effects of crossmodal associations, the mismatch condition represented a control with respect to the contingent effects observed in the match condition. The conjunction analysis between the mismatch and unimodal conditions revealed a main activation in a region along the inferior frontal sulcus of the left

hemisphere, comprising parts of Brodmann areas 44 and 9. Since the prefrontal cortex receives input from multiple sensory modalities, crossmodal effects may be found in this region. However, this area did not exhibit any effect for the match condition, a fact that suggests a different interpretation of these results: instead of reflecting the binding of information derived from different senses, this activation may be the result of the greater demand of semantic processing, required in the mismatch condition (Bentin et al. 1999).

The dorsal subdivision of area 44 of the left hemisphere has been frequently associated with the functioning of a semantic working memory system which is involved in the selection of different semantic characteristics (Martin and Chao 2001; Thompson-Schill 2003). During the mismatch condition, two incongruent stimuli were presented. Thus participants had to deal with the recognition of two different environmental events, instead of the recognition of two attributes of the same event, as in the match condition. The load on the semantic working memory system should plausibly have increased only in the mismatch condition, resulting in increased prefrontal activation.

Conclusion

Neuroimaging investigations of crossmodal processing have usually directed their attention to the importance of temporal and spatial congruence for the establishment of crossmodal associations. Our study was based on results obtained in a behavioural study about the influence of simultaneous crossmodal cues on the recognition of environmental visual and auditory stimuli. We attempted to isolate the effects of semantic congruence during a crossmodal recognition task by means of fMRI. Two crossmodal conditions were contrasted with unimodal conditions, by varying the semantic correspondence between pairs of environmental stimuli. The aim was to identify brain regions in which the crossmodal presentations exhibited a greater BOLD signal increase than in the case of single modality presentations. Effects of the congruent crossmodal presentations were primarily found in the left parahippocampal gyrus, and bilaterally in the visual unimodal cortices. These results suggest that medial temporal structures mediate the reactivation of semantic crossmodal associations, probably in association with parts of unimodal cortices. In the unimodal cortices, however, the difference between match and mismatch presentations was less significant. On the contrary, effects of incongruent crossmodal presentation were mainly found in the left inferior frontal sulcus. This result may reflect a greater demand on the semantic working memory system. Taken together, these findings suggest that it is possible to highlight crossmodal effects, due to the semantic correspondence between pairs of stimuli, in structures previously related to encoding and retrieval processes. Moreover, this investigation indicates that congruent

and incongruent crossmodal conditions elicit different cortical activations.

References

- Beauchamp MS, Lee KE, Argall BD, Martin A (2004) Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809–823
- Bentin S, Mouchetant-Rostaing Y, Giard MH, Echallier JF, Pernier J (1999) ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *J Cogn Neurosci* 11:235–260
- Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207–4241
- Bunsey M, Eichenbaum H (1996) Conservation of hippocampal memory function in rats and humans. *Nature* 379:255–257
- Calvert GA, Brammer MJ, Bullmore E, Campbell R, Iversen SD, David A (1999) Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport* 10:2619–2623
- Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in human heteromodal cortex. *Curr Biol* 10:649–657
- Calvert GA (2001) Crossmodal processing in the human brain: insight from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123
- Cox RW (1996) AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173
- Delogu F, Auricchio L, Olivetti Belardinelli M (2004) Interazioni crossmodali semantiche nell'elaborazione del suono ambientale. *Ricerche di Psicologia* (submitted)
- Driver J, Spence C (2000) Multisensory perception: Beyond modularity and convergence. *Curr Biol* 10:R731–R735
- Eichenbaum H (1992) The hippocampal system and declarative memory in animals. *J Cogn Neurosci* 4:217–231
- Ettlinger G, Wilson WA (1990) Cross-modal performance: Behavioural processes, phylogenetic considerations and neural mechanisms. *Behav Brain Res* 40:169–192
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA, Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. *Magn Reson Med* 33:636–647
- Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ (1995) Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp* 2:189–210
- Giard MH, Peronnet F (1999) Auditory-visual integration during multimodal object recognition in humans: A behavioural and electrophysiological study. *J Cogn Neurosci* 11:473–490
- Gottfried JA, Dolan RJ (2003) The nose smells what the eye sees: Crossmodal visual facilitation of human olfactory perception. *Neuron* 39:375–386
- Hadjikhani N, Roland PE (1998) Cross-modal transfer of information between the tactile and the visual representation in the human brain: A positron emission tomographic study. *J Neurosci* 18:1072–1084
- Henke K, Buck A, Weber B, Wieser HG (1997) Human hippocampus establishes associations in memory. *Hippocampus* 7(3):249–256
- Jackson O III, Schacter DL (2004) Encoding activity in anterior medial temporal lobe supports subsequent associative recognition. *Neuroimage* 21:456–462
- King AJ, Calvert G (2001) Multisensory integration: Perceptual grouping by eye and ear. *Curr Biol* 11:R322–325
- Macaluso E, Frith CD, Driver J (2000) Modulation of human visual cortex by crossmodal spatial attention. *Science* 289:1206–1208
- Martin A, Chao LL (2001) Semantic memory and the brain: Structure and processes. *Curr Opin Neurobiol* 11:194–201

- McClelland JL (1995) Constructive memory and memory distortions: A parallel-distributed processing approach. In: Schacter DL, Coyle JT, Fischbach GD, Mesulam M-M, Sullivan LE (eds) *Memory distortion*. Harvard University Press, Cambridge pp 69–90
- Mesulam M-M (1990) Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann Neurol* 28:597–613
- Mesulam M-M (1998) From sensation to cognition. *Brain* 121:1013–1052
- Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ (2002) Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. *Cogn Brain* 14:115–128
- Nadel L, Moscovitch M (1997) Memory consolidation, retrograde amnesia and the hippocampal complex. *Curr Opin Neurobiol* 7:217–227
- Price CJ, Friston KJ (1997) Cognitive conjunction: A new approach to brain activation experiments. *Neuroimage* 5:261–270
- Rombouts SAR, Machielsen WCM, Witter MP, Barkhof F, Lindeboom J (1997) Visual association encoding activates the medial temporal lobe: A functional magnetic resonance imaging study. *Hippocampus* 7(6):594–601
- Spence C, Driver J (2000) Attracting attention to the illusory location of a sound: Reflexive crossmodal orienting and ventriloquism. *Neuroreport* 11:2057–2061
- Squire LR, Zola SM (1996) Structure and function of declarative and nondeclarative memory systems. *P NAS US* 93:13515–13522
- Stark CE, Squire LR (2000) fMRI activity in the medial temporal lobe during recognition memory as a function of study-test interval. *Hippocampus* 10:329–337
- Stein BE, Meredith MA (1993) *The merging of the senses*. MIT Press, Cambridge
- Talairach J, Tournoux P (1988) *Co-planar stereotaxic atlas of the human brain*. Thieme, Stuttgart
- Thompson-Schill SL (2003) Neuroimaging studies of semantic memory: Inferring “how” from “where”. *Neuropsychology* 41:280–292