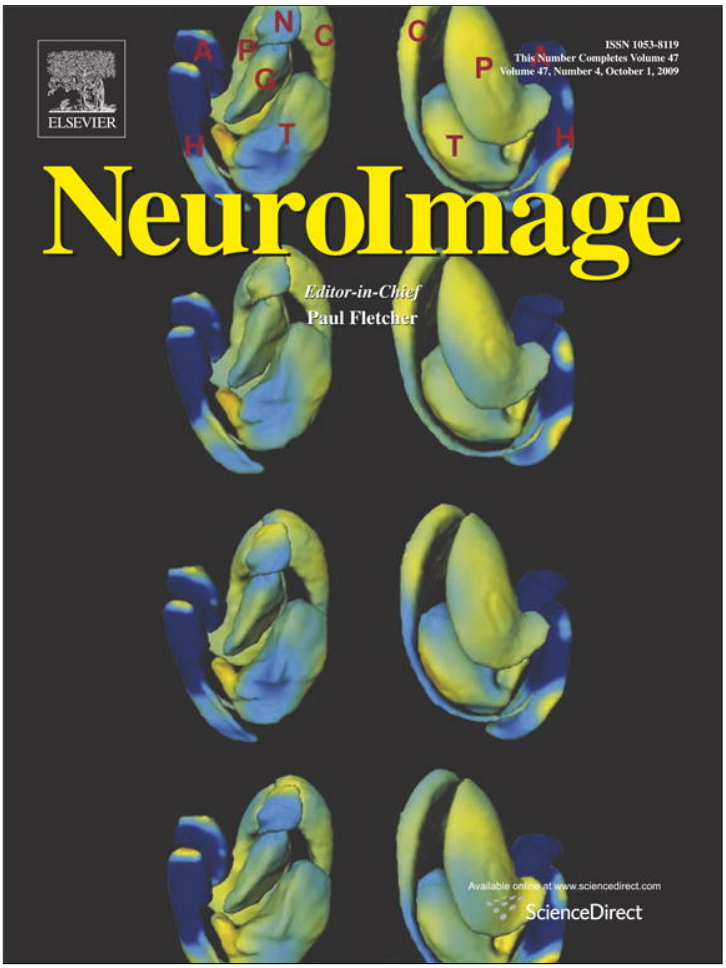

NeuroImage 47 (2009) 1992–2004

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language

Roel M. Willems [a,*,1], Aslı Özyürek [b,c], Peter Hagoort [a,c]

[a] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands
[b] Centre for Language Studies, Department of Linguistics, Radboud University Nijmegen, The Netherlands
[c] Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

ARTICLE INFO

Article history:
Received 16 November 2008
Revised 13 May 2009
Accepted 21 May 2009
Available online 1 June 2009

Keywords:
Multimodal integration
Action
Language
Gestures
Superior temporal sulcus
Inferior frontal gyrus
Semantics
Pantomimes
fMRI

ABSTRACT

Several studies indicate that both posterior superior temporal sulcus/middle temporal gyrus (pSTS/MTG) and left inferior frontal gyrus (LIFG) are involved in integrating information from different modalities. Here we investigated the respective roles of these two areas in integration of action and language information. We exploited the fact that the semantic relationship between language and different forms of action (i.e. co-speech gestures and pantomimes) is radically different. Speech and co-speech gestures are always produced together, and gestures are not unambiguously understood without speech. On the contrary, pantomimes are not necessarily produced together with speech and can be easily understood without speech. We presented speech together with these two types of communicative hand actions in matching or mismatching combinations to manipulate semantic integration load. Left and right pSTS/MTG were only involved in semantic integration of speech and pantomimes. Left IFG on the other hand was involved in integration of speech and co-speech gestures as well as of speech and pantomimes. Effective connectivity analyses showed that depending upon the semantic relationship between language and action, LIFG modulates activation levels in left pSTS.

This suggests that integration in pSTS/MTG involves the matching of two input streams for which there is a relatively stable common object representation, whereas integration in LIFG is better characterized as the on-line construction of a new and unified representation of the input streams. In conclusion, pSTS/MTG and LIFG are differentially involved in multimodal integration, crucially depending upon the semantic relationship between the input streams.

Introduction

How information streams from different modalities are integrated in the brain is a long-standing issue in (cognitive) neuroscience (e.g. Stein et al., 2004). Several studies report posterior STS/MTG as an important multimodal integration site (e.g. Calvert, 2001; Beauchamp et al., 2004a; Beauchamp et al., 2004b; Callan et al., 2004; Calvert and Thesen, 2004; van Atteveldt et al., 2004, 2007; Amedi et al., 2005; Hein and Knight, 2008). Recently, however, LIFG has been proposed to also play a role in multimodal integration when the congruency and novelty of picture and sound was modulated (Hein et al., 2007; Naumer et al., 2008), as well as in integration of information from co-speech gestures into a speech context (Willems et al., 2007). Together these studies suggest that the semantic relationship between multimodal input streams might be a relevant factor in the way different areas are recruited during multimodal integration.

Here we assessed the respective functional roles of these areas in multimodal integration by investigating responses to different language–action combinations. We exploited the fact that the semantic relation between language and action information can be rather different. Speech and co-speech gestures[2] naturally co-occur during language production and both influence the understanding of a speaker's message (e.g. McNeill, 1992, 2000; Goldin Meadow, 2003; Kita and Özyürek, 2003; Özyürek et al., 2007). For example, a speaker can move his hand laterally as he says: "The man passed by". The tight interrelatedness of speech and iconic co-speech gestures – 'gestures'

* Corresponding author.
E-mail address: roelwillems@berkeley.edu (R.M. Willems).
[1] Present address: Helen Wills Neuroscience Institute/Department of Psychology, University of California, Berkeley CA, USA.

[2] Speakers use different types of gestures as they speak (McNeill 1992; Kendon 2004; McNeill 2005). Generally speaking these could be emblems, pointing gestures, beats, or iconic gestures. In *emblems* the relations between the form and the meaning is arbitrary and emblems can be understood even in the absence of speech (i.e., an OK, 'thumbs up' gesture). In *points*, the referent can be disambiguated by the indexical relations between referent pointed at and the accompanying word (i.e., 'this pencil' pointing at pencil). *Beats* are repetitive hand movements that do not have a distinct form or meaning but co-occur with discourse or intonation breaks in the speech signal. Finally, in iconic gestures there is an iconic relation between the gesture form and the entities and events depicted. In this paper we focus on *iconic* gestures and their relation to the language. Thus we will be using the term co-speech gestures or simply 'gestures' to refer to iconic gestures for the purposes of this paper.

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.05.066

from now on – is reflected in the fact that they are hard to interpret when presented without speech (Feyereisen et al., 1988; Krauss et al., 1991; Beattie and Shovelton, 2002). Note that this does not mean that gestures are 'meaningless'. On the contrary, previous research has shown that gestures can influence understanding of a message (e.g. McNeill et al., 1994; Beattie and Shovelton, 2002; Goldin Meadow, 2003) and that they are produced for the intended addressee (Özyürek, 2002). However, gestures 'need' language to be understood, since, when they are presented without language they are not recognized unambiguously (Feyereisen et al., 1988; Krauss et al., 1991; Beattie and Shovelton, 2002). This is not true for all hand actions: pantomimes (i.e. enactions or demonstrations of an action without using an object) are produced and meant to be understood without accompanying speech (e.g. Goldin Meadow et al., 1996)[3]. Thus there is a marked difference in semantic relationship between language and gesture as compared to language and pantomimes. Gestures 'need' language to be meaningfully interpreted, whereas pantomimes can 'stand on their own' in conveying information.

The nature of neural multimodal integration may crucially depend on this difference in semantic relationship between language and action information. That is, integration of Speech–Pantomime combinations can be achieved by matching the content of two information streams onto one pre-existing representation (e.g. the verb 'stir' co-occurring with a 'stir' pantomime). However, Speech–Gesture combinations may require unifying the two streams of information into a newly constructed representation (e.g. the phrase 'The man passed by' co-occurring with a gesturing hand moving laterally; see Hagoort 2005b; Hagoort et al., in press). Previous literature indeed suggests that pSTS/MTG is more involved in integration when there is a stable common representation for the input streams (Amedi et al., 2005), whereas LIFG may be more involved in integration of novel combinations (Hein et al., 2007; Naumer et al., 2008). Here we directly assessed the functional roles of left and right pSTS/MTG and LIFG in these two types of multimodal integration. Besides changes in activation levels we investigated interactions between areas through effective connectivity analysis.

Semantic integration of language and gesture

Several neuroimaging studies have investigated the integration of semantic information conveyed through spoken language and through gestures (see Willems and Hagoort, 2007 for review). Kircher et al. (2009) observed increased activation in pSTS bilaterally, as well as in LIFG to the bimodal presentation of speech and gesture as compared to speech alone or gesture alone. Straube et al. (2009) found that better memory for methaphoric Speech–Gesture combinations was correlated with activation levels in LIFG and in middle temporal gyrus. This was interpreted as indicating better semantic integration of the two input streams, leading to higher post-test memory performance. In a related study, the integration of so-called 'beat' gestures with language was investigated. Beat gestures are supportive, rhythmic hand movements that support speech but have no semantic relationship with the speech (McNeill, 1992). Hubbard et al. found that speech combined with beats led to increased activation levels in bilateral non-primary auditory cortex, as well as in left superior temporal sulcus, as compared to speech combined with nonsense hand movements (Hubbard et al., 2009). Holle et al. (2008) presented short movie clips to participants in which an iconic gesture could disambiguate an otherwise ambiguous homonym occurring later in the sentence. The main result was that left pSTS was more strongly activated when gestures could disambiguate a homonym produced later in the sentence, as compared to meaningless 'grooming' movements. Finally, in an earlier report we employed a semantic mismatch paradigm to investigate semantic integration of speech and co-speech iconic gestures (Willems et al., 2007). An increase in activation level in LIFG was observed, both during semantic integration of gestures as well as during semantic integration of speech.

These studies show that LIFG and pSTS/MTG which are thought to be implicated in multimodal integration, are also active during integration of language and gestures. However, there is a marked discrepancy between whether both LIFG and pSTS/MTG, or only one of the two is found active during Speech–Gesture integration. It is viable that these differences are due to differences in stimulus materials, which ranged from relatively abstract beat gestures (Hubbard et al., 2009), to metaphoric gestures (Straube et al., 2009; Kircher et al., 2009), to iconic gestures (Willems et al., 2007; Holle et al., 2008). In the present paper we want to get a better insight into the respective roles of pSTS/MTG and LIFG during integration of language and action information.

Present study

As stated above, our main goal was to see whether and how the semantic relationship between input streams would change the involvement of multimodal integration areas. Participants were presented with unimodal gesture/pantomime videos and audio content, as well as bimodal Speech–Gesture and Speech–Pantomime combinations. In the bimodal conditions speech and action content could either be in accordance or in discordance with each other. We choose the congruency paradigm because it has been shown that semantically discordant Speech–Gesture combinations successfully increase semantic processing load (Willems et al., 2007; see also Willems et al., 2008b). Moreover, studies that increase semantic processing load without using semantically incongruent stimuli find similar neural correlates than studies employing a mismatch paradigm (see Hagoort et al., 2004; Rodd et al., 2005; Davis et al., 2007). That is, by using this paradigm we assessed whether a multimodal area is involved in integrating the two streams of information at the semantic level (Beauchamp et al., 2004b; Hein et al., 2007; Hocking and Price, 2008).

A recent investigation suggests that comparing incongruent to congruent multimodal combinations is a useful additional test for multimodal integration next to comparing a bimodal response to the combination of unimodal responses (Hocking and Price, 2008). That is, Hocking and Price showed that pSTS/MTG exhibits a similar response to integration of audio–visual stimulus pairs (e.g. the spoken word 'guitar' presented after the picture of a guitar) as to audio–audio (e.g. the spoken word 'guitar' presented after the sound of a guitar) or visuo–visuo pairs (e.g. the written word 'guitar' presented after a picture of a guitar). The authors argued that these data show that pSTS/MTG is not so much involved in combining information from different input channels, since it is equally activated when the input format is the same (e.g. spoken word 'guitar' presented after sound of a guitar). Rather, they propose that pSTS/MTG's function is that of conceptual matching, irrespective of input modality (Hocking and Price, 2008). They observed that pSTS was sensitive to a congruency manipulation in the bimodal input. Hence, we presented our stimuli in unimodal presentation formats, as well as in bimodal congruent and in bimodal incongruent format.

Previous literature hints at the suggestion that pSTS/MTG is more involved in integration when there is a stable common representation for the input streams (Amedi et al., 2005) as compared to LIFG which may be more involved in the on-line creation of a novel representation (Hein et al., 2007; Naumer et al., 2008). Thus we expect to see differential involvement of pSTS/MTG for Speech–Pantomime combinations, in which the two input streams can be mapped onto a

[3] Note that even though pantomimic gestures can also be used accompanying speech for demonstration purposes when speakers quote their own or others' actions (Clark and Gerrig 1990), they do not have to be. In fact speakers usually interrupt speaking for a while for pantomimic demonstrations, whereas iconic gestures are used 90% of the time during speaking (McNeill 1992).

relatively stable representation in long-term memory. On the contrary, LIFG may be sensitive to Speech–Gesture combinations since in this case a novel representation needs to be established. If these predictions are correct, we should only observe differences in pSTS/MTG depending upon congruency in Speech–Pantomime combinations, but not to congruency in Speech–Gesture combinations. On the contrary, if LIFG is involved in on-line unification of information when a novel representation has to be established, we expect to see different activation levels in this area to congruency in Speech–Gesture combinations. Moreover, given the well-known modulatory function of frontal cortex (Miller, 2000; Gazzaley and D'Esposito, 2007), it is conceivable that LIFG modulates other areas during multimodal integration (such as pSTS/MTG). We tested this by means of effective connectivity analysis (Friston et al., 1997).

Materials and methods

Participants

Twenty healthy right-handed (Oldfield, 1971) participants without hearing complaints and with normal or corrected-to-normal vision took part in the experiment. None of the participants had any known neurological history. Data of four participants were not analyzed because they did not perform significantly above chance level. Data from the remaining 16 participants (11 female; mean age = 22.3 years, range = 19.3–27.4 years) were entered into the analysis. The study was approved by the local ethics committee and all participants gave informed consent prior to the experiment in accordance with the Declaration of Helsinki.

Materials

Stimuli consisted of Speech–Gesture segments or Speech–Pantomime combinations. These were presented either in matching (Gest-Match, Pant-Match) or in mismatching (Gest-Mism, Pant-Mism) combinations of gestures/pantomimes with speech. The label 'Match'/'Mism' refers to the match of gesture/pantomime with speech. In the unimodal runs (see below) the video or audio content of all segments was presented. Below we first describe how stimuli were selected and then turn to the experimental design.

Iconic gestures (i.e., gestures about actions and/or objects) (McNeill, 1992) were taken from a natural retelling of cartoon movies by a female native speaker of Dutch (Fig. 1A and Appendix A). For the pantomimes we asked another female native speaker of Dutch to pantomime common actions, i.e. to enact an action without using the object normally associated with that action (Fig. 1B and Appendix B). All videos were recorded in a sound-shielded room with a Sony TCR-TRV950 PAL camera. The actor's head was kept out of view to eliminate influences of lip or head movements. Short segments of Speech–Gesture combinations were cut from the overall retelling using Adobe Premier Pro software (version 7.0; www.adobe.com). All gesture segments contained one or more gestures with iconic content, such as referring to motion events or actions (see the Appendix A for a literal transcription of the materials). The original audio content from the natural retelling was taken for the gesture materials. The audio content of the pantomimes (i.e. spoken verbs) was re-recorded after recording of the video and were spoken by the same actor as in the videos. All audio content was band-pass filtered from 80 to 10500 Hz and equalized in sound level to 80 dB using 'Praat' software (version 4.3.16; www.praat.org). Finally, the speech files were edited into the video files to create matching or mismatching Speech–Gesture or Speech–Pantomime combinations.

Stimuli were selected on the basis of two pretests. In pretest 1, naive raters ($n = 20$, not participating in fMRI session) had to indicate what they thought was being depicted in the gesture/pantomime videos (presented without speech). In pretest 2, a group of different raters ($n = 16$, not participating in fMRI session) judged how well speech and gesture or speech and pantomime combinations matched on a 1–5 scale (results below). The final stimulus set used in the fMRI session contained 12 matching Speech–Gesture combinations and 12 matching Speech–Pantomime combinations, as well as an equal amount of Speech–Gesture and Speech–Pantomime mismatches.

The results from the two pretests for the final set of stimuli are described below and are summarized in Table 1. The meaning of the 12 co-speech gestures was not easily recognizable without speech (results pretest 1, mean percentage of raters ($n = 20$) that indicated the correct meaning to a gesture: 8.8%, standard deviation (s.d.) = 13.7%). On the other hand, the meaning of the 12 pantomimes was highly recognizable without speech (pretest 1, mean percentage of raters ($n = 20$) that assigned the correct meaning to a pantomime: 88.4%, s.d. = 14.7%). The results of pretest 2 showed that the original combinations of gesture and speech were scored as matching whereas the mismatching pairs were scored as mismatching (results pretest 2: matching: mean = 3.90, s.d. = 0.64; mismatching: mean = 1.74, s.d. = 0.49, on a 1–5 scale). Similarly for pantomimes and speech, the matching combinations were consistently recognized as matching,

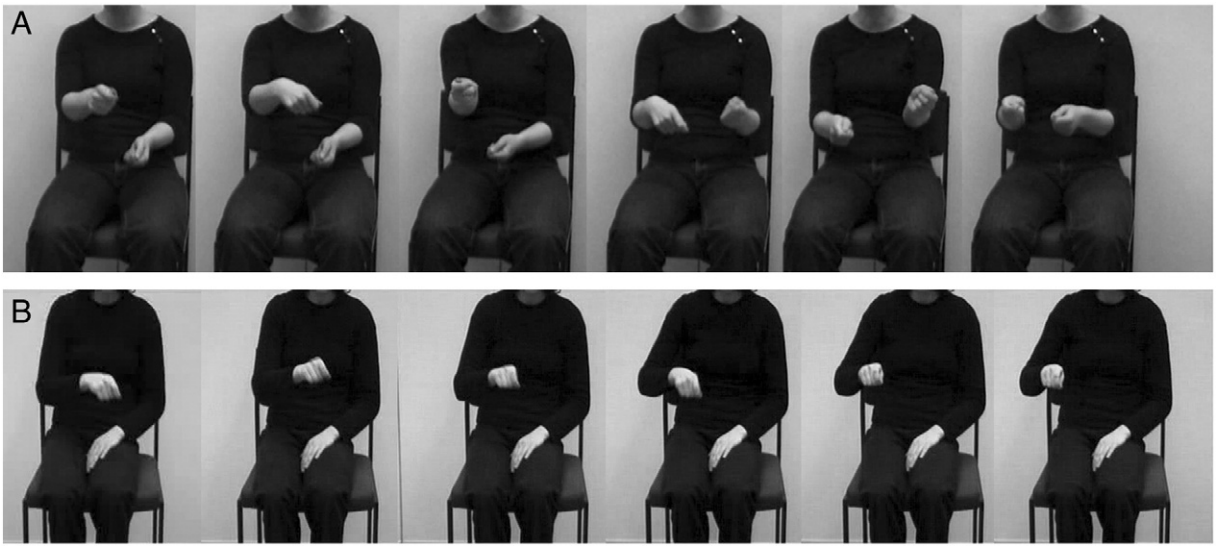

Fig. 1. Examples of video content of the stimulus materials. (A) Six stills of one of the gestures. This gesture is taken from a segment in which the speaker describes a character writing and drawing on a paper on a table. For exact speech see Appendix A. (B) Six stills of one the pantomimes ('to write'). Materials were presented in color.

Table 1
Characteristics of stimuli.

Stimulus type	Pretest 1		Pretest 2	
	Mean (% participants)	s.d.	Mean (score)	s.d.
Pantomimes	88.4	14.7		
Gestures	8.8	13.7		
Pant-Speech match			4.95	0.07
Pant-Speech mismatch			1.09	0.13
Gest-Speech match			3.90	0.64
Gest-Speech mismatch			1.74	0.49

Table shows the results of two pretests for the final set of stimuli. Pretest 1 involved presenting pantomimes and gestures without speech and asking raters (different than the participants who took part in the fMRI session) to indicate what they thought was depicted in the actions. Displayed is the mean percentage of raters ($n=20$) that indicated the meaning that matched the meaning of the pantomime or the meaning in the original speech fragment (for gestures). In pretest 2, matching and mismatching Speech–Pantomime and Speech–Gesture combinations were presented. A new group of raters ($n=16$, different than the participants who took part in the fMRI session) had to indicate how well they thought audio and video fit together on a 1–5 point scale. The final stimuli were selected to ensure that the meaning of the Pantomimes, but not of the Gestures, was reliably recognizable when presented without speech (pretest 1) and that Matching and Mismatching combinations would be perceived as such (pretest 2).

whereas the mismatching combinations were not (matching: mean = 4.95, s.d. = 0.07; mismatching: mean = 1.09, s.d. = 0.13, on a 1–5 scale). Despite the differences in spread (see standard deviations), scores for matching and mismatching Speech–Gesture combinations were not different from matching and mismatching Speech–Pantomime combinations ($t(23)=-1.01$, $p=0.33$). Nevertheless, we took the difference in spread in these congruency scores into account in the fMRI data analysis (see below).

Mean duration of the stimuli was 2028 ms (s.d. = 506; range = 1166–3481) for the Pantomimes and 2209 ms (s.d. = 400; range = 1366–3182) for the Gestures. Note that in the main analysis we did not directly compare Pantomimes and Gestures given that these stimulus sets were not matched on basic characteristics such as duration.

Experimental procedure

There were three experimental runs: audio with video (AV), audio only (AUDIO), and video only (VIDEO). The unimodal runs were included to test whether integration areas were also activated during unimodal presentation of the stimuli.

In the AV run participants saw the Speech–Gesture and Speech–Pantomime combinations, in matching and in mismatching versions. The 12 matching and 12 mismatching combinations were repeated three times each, leading to 36 trials per condition (Gest-Match, Gest-Mism, Pant-Match, Pant-Mism). There were 4 matching and 4 mismatching filler items (taken from the materials that were rejected based on the pretests) for both Speech–Gesture and Speech–Pantomime combinations. These were all repeated two times, leading to a total of 32 filler trials (16 gesture, 16 pantomime). Filler items were included to ensure participants were paying attention to the stimuli (see below).

In the AUDIO run participants heard the short utterances or verbs from the gesture and pantomime recordings without visual content on the screen. There were 12 pantomime and 12 gesture audio stimuli, which were all repeated three times, leading to 36 trials for each condition (Gest-Audio, Pant-Audio). In the VIDEO run participants saw the gestures and pantomimes presented without speech. Again, there were 12 gesture and 12 pantomime stimuli which were repeated three times leading to 36 trials per condition (Gest-Video, Pant-Video). In both the AUDIO and the VIDEO runs, eight filler stimuli were presented, four gestures and four pantomimes. Fillers were repeated two times, leading to a total of 16 filler trials (8 gesture, 8 pantomime).

Stimuli were presented using 'Presentation' software (version 10.2; www.nbs.com). The visual content was displayed from outside of the scanner room onto a mirror above the participant's eyes, mounted onto the head coil. The auditory content was presented through sound reducing MR-compatible head phones. The sound level was adjusted to the preference of each participant during a practice run in which ten items, which were not used in the remainder of the experiment, were presented while the scanner was switched on. All participants indicated that they could hear the auditory stimuli well, and none of the participants asked for the sound level to be increased to more than its half-maximum.

After each filler item (22% of the trials), a screen was presented with 'yes' and 'no' on either the left or the right side of the screen. Participants had to indicate whether they had observed that specific stimulus item before or not, by pressing a button with either the left or the right index finger. Response side was balanced over filler trials such that 'yes' was indicated with the left index finger in one half of the filler trials and with the right index finger in the other half of the filler trials. Participants had 2.5 s to respond and were instructed to respond as accurately as possible. Feedback was given after each response by appearance of the word 'correct', 'incorrect' or 'too late' on the screen. This task was employed to ensure that participants would be actively processing the stimuli.

Stimuli were presented in an event-related fashion, with an average intertrial interval (ITI) of 3.5 s. Onset of the stimuli was effectively jittered with respect to volume acquisition by varying the ITI between 2.5 and 4.5 s in steps of 250 ms (Dale, 1999). The order of conditions was pseudo-randomized with the constraint that a condition never occurred three times in a row. Four stimulus lists were created which were evenly distributed over participants. The order of runs was varied across participants.

Image acquisition

Data acquisition was performed using a Siemens 'Trio' MR-scanner with 3 T magnetic field strength. Whole-brain echo-planar images (EPIs) were acquired using a bird-cage head coil with single pulse excitation with ascending slice order (TR = 2130 ms, TE = 30 ms, flip angle = 80 degrees, 32 slices, slice thickness = 3 mm, 0.5 mm gap between slices, voxel size $3.5\times3.5\times3$ mm). A high resolution T1 weighted scan was acquired for each subject after the functional runs using an MPRAGE sequence (192 slices, TR = 2300 ms; TE = 3.93 ms; slice thickness = 1 mm; voxel size $1\times1\times1$ mm).

Data analysis

Data were analyzed using SPM5 (http://www.fil.ion.ucl.ac.uk/spm/software/spm5/). Preprocessing involved discarding the first four volumes, correction of slice acquisition time to time of acquisition of the first slice, motion correction by means of rigid body registration along 3 rotations and 3 translations, normalization to a standard MNI EPI template including interpolation to $2\times2\times2$ mm voxel sizes, high-pass filtering (time constant of 128 s) and spatial smoothing with an 8 mm FWHM Gaussian kernel. Statistical analysis was performed in the context of the General Linear Model (GLM) with regressors 'Gestures' and 'Pantomimes' in the AUDIO and VIDEO runs and regressors 'Gest-Match', 'Gest-Mism', 'Pant-Match', 'Pant-Mism' in the AV run. Additionally, responses (i.e. button presses), filler items and the motion parameters from the motion correction algorithm were included in the model. All regressors except for the motion parameters were convolved with a canonical two-gamma hemodynamic response function. Visualization of statistical maps was done using MRIcroN software (http://www.sph.sc.edu/comd/rorden/mricron/).

As explained in the Introduction we had an a priori hypothesis that LIFG and pSTS/MTG would be involved in integration of action and language information. Therefore we created regions of interest (ROIs) in these areas. For LIFG we took the mean of the maxima from inferior frontal cortex from a recent extensive meta-analysis of neuroimaging studies of semantic processing (Vigneau et al., 2006) (centre coordinate: MNI [−42 19 14]). The ROIs in left and right pSTS/MTG were based upon

a recent meta-analysis of multimodal integration studies (Hein and Knight, 2008) (centre coordinate left: MNI [−49 −55 14]; right: MNI [50 −49 13]). Regions of interest were spheres with an 8 mm radius. The activation levels of all voxels in a ROI were averaged for each subject separately and differences between conditions were assessed by means of dependent samples *t*-test with $df = 15$.

We subsequently and additionally tested whether there was a relationship between the *degree* of congruence between speech and gesture or speech and pantomime and activation levels in these two ROIs. The scores from pretest 2 (in which raters indicated how well they thought action and speech were in accordance with each other, see Table 1) show that all Speech–Pantomime combinations were judged as clearly matching (mean on 1–5 point scale = 4.95, s.d. = 0.07) or mismatching (mean = 1.09, s.d. = 0.13). However, in the Speech–Gesture pairs there was considerably more spread in these scores, both in the matching combinations (mean = 3.90, s.d. = 0.64) as well as in the mismatching combinations (mean = 1.74, s.d. = 0.49). Therefore we reasoned that by using a parametrically varying regressor based upon these scores, we would be able to pick up effects of Speech–Gesture congruence in a more sensitive way than by comparing all mismatching Speech–Gesture combinations to all matching Speech–Gesture combinations. For each stimulus item, the mean score (ranging from 1 to 5) from the pretest was taken and a linearly varying parametric regressor was constructed (Buchel et al., 1998). It should be noted that these scores were obtained from a different group of participants (raters) as that participated in the fMRI experiment and that hence, the perceived congruence between Speech–Gesture/Speech–Pantomime combinations may be different in our fMRI participants. Although this cannot be ruled out, it is not very likely given that the raters were recruited from the same population as the fMRI participants (i.e. Nijmegen undergraduates) and that the ratings were taken from a reasonable amount of raters ($n = 16$).

A whole-brain analysis was performed by taking single subject contrast maps to the group level with factor 'subjects' as a random factor (random effects analysis). In two separate analyses areas were determined that responded more strongly to bimodal as compared to unimodal stimulus presentation. First we investigated which areas were more strongly activated to bimodal presentation as compared to each unimodal condition in isolation, and which responded above baseline in the unimodal conditions. This was implemented as a conjunction analysis (testing a logical AND, Nichols et al., 2005) of the combined bimodal condition to audio-alone and to video-alone (i.e. comparisons Pant-match + Pant-mismatch > Pant-audio ∩ Pant-match + Pant-mismatch > Pant-video; and Gest-match + Gest-mismatch > Gest-audio ∩ Gest-match + Gest-mismatch > Gest-video). Each comparison was inclusively masked with the conjunction of the unimodal conditions compared to zero (i.e. Pant-video > 0 ∩ Pant-audio > 0 and Gest-video > 0 ∩ Gest-audio > 0). Contrasts were balanced by weighting the unimodal conditions twice as strongly as the bimodal conditions. Second we investigated which areas responded more strongly to the combined bimodal conditions as compared to the combined unimodal conditions (Pant-match + Pant-mism > Pant-audio + Pant-video; and Gest-match + Gest-mism > Gest-audio + Gest-video).

Table 2
Response characteristics of the a priori defined ROIs during uni- and bimodal presentation of the stimuli.

Region	AUDIO > 0				VIDEO > 0			
	Pant $t(15)$	p	Gest $t(15)$	p	Pant $t(15)$	p	Gest $t(15)$	p
Left pSTS/MTG	3.71	**0.001**	3.67	**0.001**	1.84	**0.043**	2.14	**0.025**
Right pSTS/MTG	5.15	<**0.001**	3.90	<**0.001**	2.53	**0.012**	3.23	**0.003**
LIFG	3.03	**0.004**	3.85	<**0.001**	3.31	**0.002**	4.14	<**0.001**
	AV > A				AV > V			
	Pant $t(15)$	p	Gest $t(15)$	p	Pant $t(15)$	p	Gest $t(15)$	p
Left pSTS/MTG	3.28	**0.005**	3.33	**0.004**	4.31	<**0.001**	3.69	**0.002**
Right pSTS/MTG	4.19	<**0.001**	3.08	**0.007**	4.92	<**0.001**	3.95	**0.001**
LIFG	2.78	**0.007**	3.39	**0.002**	3.42	**0.002**	2.99	**0.005**
	Mean (AV) > Mean (A + V)							
	Pant $t(15)$	p	Gest $t(15)$	p				
Left pSTS/MTG	5.38	<**0.001**	4.70	<**0.001**				
Right pSTS/MTG	4.99	<**0.001**	3.85	<**0.001**				
LIFG	3.79	<**0.001**	3.72	**0.001**				
	Mean (AV) > Max (A, V)							
	Pant $t(15)$	p	Gest $t(15)$	p				
Left pSTS/MTG	2.42	**0.029**	2.79	**0.013**				
Right pSTS/MTG	2.95	**0.010**	3.42	**0.004**				
LIFG	3.14	**0.007**	2.10	**0.053**				
	Match > 0							
	Pant $t(15)$	p	Gest $t(15)$	p				
Left pSTS/MTG	5.19	<**0.001**	5.10	<**0.001**				
Right pSTS/MTG	10.33	<**0.001**	12.78	<**0.001**				
LIFG	4.71	<**0.001**	5.07	**0.001**				

All ROIs (left pSTS/MTG, right pSTS/MTG and LIFG) were activated above baseline during unimodal presentation of the stimuli (first panel). Bimodal presentation of the stimuli led to higher activation levels than either audio only or video only presentation (second panel). Moreover, all ROIs are more strongly activated during bimodal presentation of the stimuli as compared to the mean of the two unimodal presentations (third panel, 'mean criterion'), as well as compared to the maximum of the unimodal conditions (fourth panel, 'max criterion') (Beauchamp, 2005b). Finally, all bimodal match conditions activated the ROIs significantly stronger than baseline (lower panel). Bold typeface indicates statistical significance at the $p < 0.05$ level.

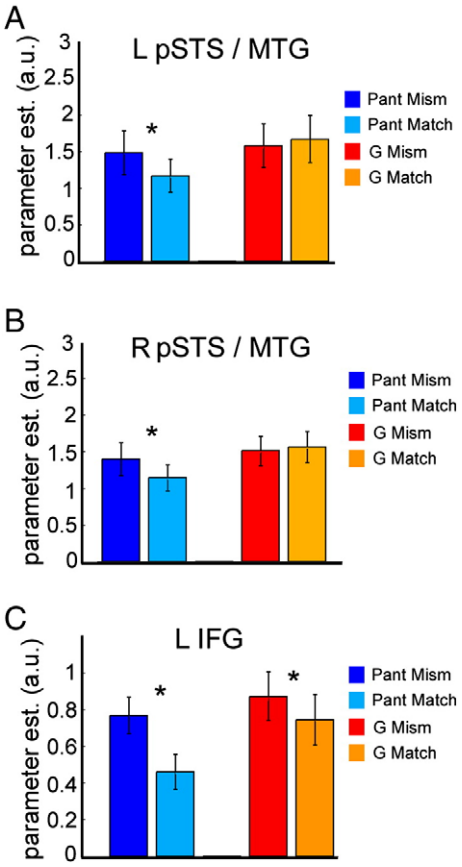

Fig. 2. Results in a priori defined Regions of Interest. Mean parameter estimates of all bimodal conditions in left pSTS/MTG (A), right pSTS/MTG (B) and LIFG (C), averaged over all voxels in the ROI. (A) In left pSTS/MTG there was a difference between mismatching and matching Speech–Pantomime combinations (mismatch: dark blue, match: light blue), but not between mismatching and matching Speech–Gesture combinations (mismatch: red; match: orange). (B) A similar pattern of responses was observed in right pSTS/MTG. (C) On the contrary, in LIFG, there was an influence of congruence both in the Speech–Pantomime combinations as well as in the Speech–Gesture combinations. Asterisks indicate significance at the $p<0.05$ level.

Whole-brain correction for multiple comparisons was applied by combining a significance level of $p=0.001$, uncorrected at the voxel level, with a cluster extent threshold using the theory of Gaussian random fields (Friston et al., 1996). All clusters are reported at an alpha level of $p<0.05$ corrected across the whole brain. Anatomical localization was done with reference to the atlas by Duvernoy (1999).

Finally we investigated effective connectivity of LIFG and pSTS/MTG onto other cortical areas by means of whole-brain Psycho-Physiological Interactions (PPIs) (Friston et al., 1997; Friston, 2002). A PPI reflects a change in the influence of one area onto other areas depending upon the experimental context. We performed two PPI analyses: one looking for effective connectivity of pSTS/MTG or LIFG (a priori defined regions of interest defined above) with other areas, modulated by Speech–Pantomime match/mismatch, and the other one looking for modulations in connectivity between each of these two areas and other areas during Speech–Gesture match/mismatch. Time courses were deconvolved with a canonical hemodynamic response function, as suggested by Gitelman et al. (2003). Again, whole-brain family-wise error correction for multiple comparisons was applied by combining a significance level of $p=0.001$, uncorrected at the voxel level, with a cluster extent threshold using the theory of Gaussian random fields (Friston et al., 1996). All clusters are reported at an alpha level of $p<0.05$ corrected across the whole brain.

Results

Behavioral results

Four participants did not score above chance level to the filler items in at least one of the runs and were discarded from further analysis. Performance of the remaining 16 participants was well above chance level indicating that participants attended the stimuli (AUDIO: mean percentage correct = 83.75, range = 64.3–93.8, s.d. = 9.26; VIDEO: mean percentage correct = 77.21, range = 62.5–92.3, s.d. = 10.27; AV: mean percentage correct = 75.42, range = 62.1–87.5, s.d. = 7.84). In a repeated measures ANOVA, with factor Run (AV, A, V), there was a marginally significant main effect of Run ($F(1, 30)=3.63$, $p=0.055$). Planned comparisons showed that the AV run was significantly more difficult than the AUDIO run ($F(1,15)=15.47$, $p=0.001$), but not than the VIDEO run ($F(1,15)<1$). The AUDIO and VIDEO run were not significantly different from each other, although there was a trend for the VIDEO run to be more difficult ($F(1,15)=3.08$, $p=0.10$).

We separately analyzed the behavioral results from the AV run (mean percentage correct: Pant-Match: 79.7% (s.d. 13.9), Pant-Mism: 75.2 (s.d. 12.9), Gest-Match: 80.4% (s.d. 11.2), Gest-Mism: 65.8% (s.d. 15.7)). All these scores were above chance level (all $p<0.001$). A repeated measures ANOVA with factors Congruency (Match, Mismatch) and Stimulus type (Pant, Gest) revealed a main effect of Congruency ($F(1, 15)=12.29$, Mse = 0.012, $p=0.003$), but no main effect of Stimulus type ($F(1,15)=1.62$, Mse = 0.019, $p=0.22$), or a Congruency × Stimulus type interaction ($F(3,45)=2.35$, Mse = 0.017, $p=0.16$), indicating that performance was not significantly different for Pantomime or Gesture stimuli and that there was no interaction effect between the congruency of the stimuli and whether they were Speech–Pantomime combinations or Speech–Gesture combinations.

Region of interest analysis

Bimodal versus unimodal presentation

All ROIs were activated above baseline in all unimodal conditions. Moreover, the bimodal conditions (collapsed over matching and mismatching combinations) led to stronger activation as compared to each unimodal condition in isolation, as well as to the mean, as well as to the maximum of the unimodal conditions (Beauchamp, 2005b) (see Table 2). That is, all ROIs fulfilled the following criteria: $AV>(A+V)/2$ ('mean criterion'), and $AV>\max(A,V)$ ('max criterion'), and $0<V<AV>A>0$.

Congruency effects

In the ROI in left pSTS/MTG, activation levels were significantly higher in the Pant-Mism as compared to Pant-Match condition ($t(15)=2.76$, $p=0.007$) (Fig. 2A, Table 3). No such effect was observed for Speech–Gesture combinations (Gest-Mism vs. Gest-Match: $t(15)=$

Table 3
Results in a priori defined regions of interest comparing Pant-Mism versus Pant-Match and Gest-Mism versus Gest-Match.

Region	Pant-Mism vs. Pant-Match		Gest-Mism vs. Gest-Match	
	$t(15)$	p	$t(15)$	p
Left pSTS/MTG	2.76	**0.007**	−1.17	n.s.
Right pSTS/MTG	2.17	**0.023**	<1	n.s.
LIFG	6.01	<**0.001**	1.75	**0.050**

Left as well as right pSTS/MTG were sensitive to congruence in Speech–Pantomime combinations, but not in Speech–Gesture combinations. However, LIFG was sensitive to congruence both in Speech–Pantomime combinations as well as in Speech–Gesture combinations. Regions of interest were 8 mm spheres around centre voxels which were taken from two meta-analyses (Vigneau et al., 2006; Hein and Knight 2008). MNI coordinates were [−42 19 14] for LIFG, and [−49 −55 14] and [50 −49 13] for left and right pSTS/MTG. Bold typeface indicates statistical significance at the $p<0.05$ level.

−1.17, n.s.). A similar pattern was observed in right pSTS/MTG (Pant-Mism vs. Pant-Match: $t(15)=2.17$, $p=0.023$; Gest-Mism vs. Gest-Match: $t(15)<1$) (Fig. 2B, Table 3). However, in LIFG, activation levels were higher both for Pant-Mism as compared to Pant-Match conditions ($t(15)=6.01$, $p<0.001$) as well as for Gest-Mism as compared to Gest-Match conditions ($t(15)=1.75$, $p=0.050$) (Fig. 2C, Table 3). Testing the degree of congruence (based upon the results of pretest 2 in which participants had to indicate how well Speech–Pantomime or Speech–Gesture combinations matched), showed a similar pattern of results. In LIFG there was an effect of degree of congruence both for Speech–Gesture combinations ($t(15)=2.39$, $p=0.015$) as well as for Speech–Pantomime combinations ($t(15)=5.58$, $p<0.001$) (Supplementary Table S1). However in left and right pSTS/MTG there was only an effect for the Speech–Pantomime combinations (left: $t(15)=4.29$, $p<0.001$; right: $t(15)=2.22$, $p=0.021$) but not for Speech–Gesture combinations (left: $t(15)<1$; right: $t(15)<1$) (Supplementary Table S1). This confirms the previous ROI analysis and rules out the possibility that the absence of an effect of Gest-Mism versus Gest-Match in left pSTS/MTG is due to the larger spread of congruence scores in the Speech–Gesture combinations.

In a separate analysis we compared the magnitude of the congruency effect for Speech–Gesture and Speech–Pantomime combinations. That is, we compared (Pant-mism>Pant-match)>(Gest-mism>Gest-match) in a two-sided t-test in each ROI separately. The results show that in all ROIs the congruency effect in the Pant-Speech combinations was larger as compared to the congruency effect in the Speech–Gesture combinations (L pSTS/MTG: $t(15)=2.93$, $p=0.010$; R pSTS/MTG: $t(15)=6.55$, $p<0.001$; LIFG: $t(15)=2.17$, $p=0.047$). However, as is clear from Fig. 2 as well as from the results described above, in LIFG this difference was relative: there was a congruency effect both in the Speech–Pantomime as well as in the Speech–Gesture combinations. This was crucially not the case in left and right pSTS/MTG. In bilateral pSTS/MTG only a congruency effect for Pant-mism>Pant-match was observed.

Although the error bars in Fig. 2 do not suggest so, it is possible that we did not find a congruency effect for Speech–Gesture combinations in pSTS/MTG because the variance in this region was different for Speech–Gesture as compared to Speech–Pantomime combinations. We tested for homogeneity of variances in all three ROIs, comparing variance in the Speech–Gesture combinations to variance in the Speech–Pantomime combinations using Levene's test (Levene, 1960). No such differences in variances were observed (all $F<1$), suggesting that the lack of a congruency effect in left and right pSTS/MTG to Speech–Gesture combinations is not due to a different amount of variance as compared to Speech–Pantomime combinations.

Whole-brain analysis

Bimodal versus unimodal presentation

We first investigated in which regions bimodal conditions elicited stronger activations than in each unimodal condition in isolation. For the Speech–Pantomime combinations increased activations were observed in bilateral pSTS, LIFG, and in bilateral inferior occipital sulcus (Fig. 3A and Table 4). For the Speech–Gesture combinations, activations were observed in the same set of regions: bilateral pSTS and in LIFG (Fig. 3B and Table 4).

Second, we looked at regions which showed a stronger activation during bimodal as compared to the sum of the unimodal conditions (Pant-match + Pant-mismatch > Pant-audio + Pant-video and Gest-match + Gest-mism > Gest-audio + Gest-video). For Speech–Pantomime combinations this lead to a wide-spread network of areas encompassing LIFG, bilateral superior temporal gyri, bilateral superior temporal sulci, bilateral planum temporale, and extensive activations in early visual areas including bilateral inferior and middle occipital gyri as well as the thalamus bilaterally (Fig. 3C and Table 4). For the Speech–Gesture combinations a highly similar pattern of activations was observed, including LIFG, bilateral superior temporal sulci/gyri, bilateral planum temporale, bilateral inferior and middle occipital gyri and the thalamus bilaterally (Fig. 3D and Table 4).

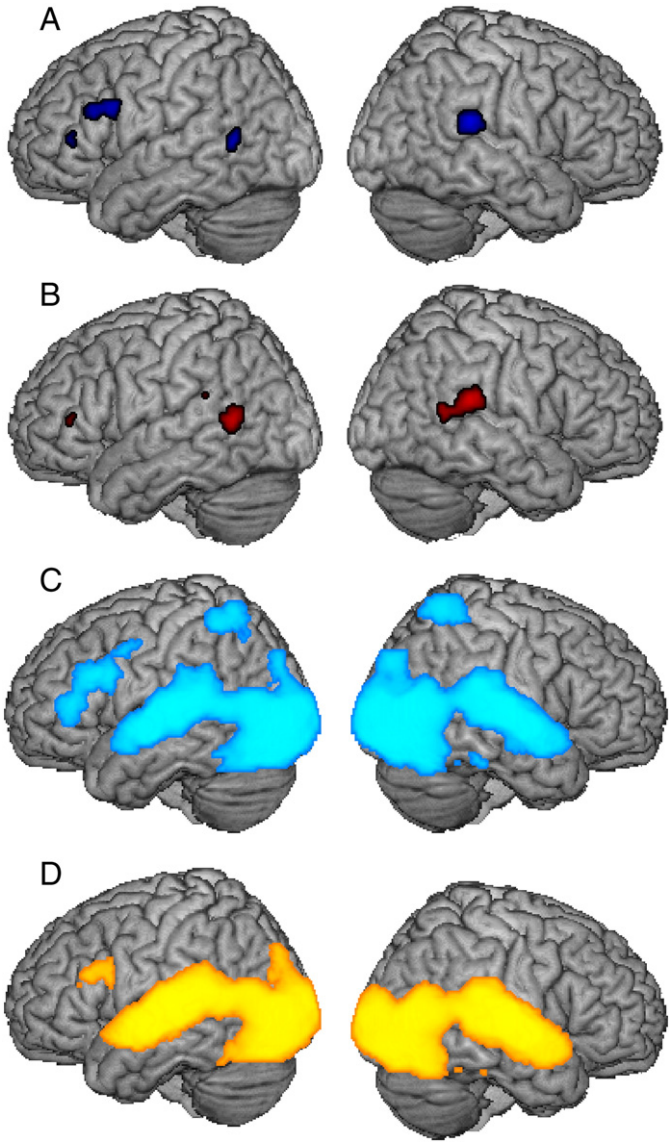

Fig. 3. Results from whole-brain analysis comparing bimodal to unimodal conditions. The upper two panels (A and B) show areas more strongly activated to bimodal stimuli as compared to audio-alone and as compared to video-alone. This was implemented as a conjunction analysis (Nichols et al., 2005) comparing Pant-match + Pant-mismatch > Pant-audio only ∩ Pant-match + Pant-mism > Pant-video only (A) or Gest-match + Gest-mism > Gest-audio only ∩ Gest-match + Gest-mism > Gest-video only (B), which was implicitly masked with a conjunction of both unimodal conditions > baseline. Contrast weights were balanced such that the unimodal was weighted twice as strong as each unimodal condition. The lower two panels (C and D) show results comparing the combined bimodal conditions to the combined unimodal conditions, that is, Pant-match + Pant-mism > Pant-audio + Pant-video (C) and Gest-match + Gest-mism > Gest-audio + Gest-video (D). Results are displayed at $p<0.05$, corrected for multiple comparisons.

Congruency effects

Contrasting Pant-Mism with Pant-Match led to a network of areas encompassing left and right pSTS/MTG, LIFG, left intraparietal sulcus, bilateral insula and bilateral cingulate sulcus (Fig. 4 and Table 5). Note that the clusters of activation in left and right pSTS/MTG overlapped with the a priori defined ROIs in left and right pSTS/MTG.

There were no areas which survived the statistical threshold to the Gest-Mism versus Gest-Match comparison. However, informal

Table 4
Results of whole-brain analyses comparing bimodal versus unimodal conditions.

Contrast	Region	T(max)	Coordinates (MNI)		
0<Pant-video<Pant-bimodal>Pant-audio>0	L pSTS	4.62	−48	−54	8
	R pSTS	5.04	62	−36	18
			58	−52	10
	L IFG	3.49	−40	10	26
0<Gest-video<Gest-bimodal>Gest-audio>0	L pSTS	4.16	−52	−56	6
	R pSTS	4.92	58	−36	18
	L IFG	4.26	−46	10	24
			−46	22	24
	R inf occipital sulcus	4.05	14	−92	−2
			4	−92	−2
	L inf occipital sulcus	3.60	−24	−76	−10
Pant-bimodal>Pant-video + Pant-audio	R superior temporal gyrus/sulcus	7.06	60	−32	12
			50	−12	−2
	L superior temporal gyrus	5.58	−48	−38	14
			−52	−12	−4
	R planum temporale	6.79	−60	−38	17
	L planum temporale	5.56	60	−32	16
	LIFG	5.44	−46	12	26
	Left middle occipital gyrus	7.11	−44	−82	4
	Left inferior occipital gyrus	4.95	−8	−104	0
	Right middle occipital gyrus	8.67	48	−68	4
	Right inferior occipital gyrus	9.07	20	−96	−8
	Left thalamus	6.16	−16	−32	−2
	Right thalamus	5.91	22	−28	−2
Gest-bimodal>Gest-video + Gest-audio	R superior temporal gyrus/sulcus	11.60	60	−18	2
			63	−31	15
	L superior temporal gyrus/sulcus	9.47	−66	−30	10
			−54	−32	11
	R planum temporale	7.06	60	−32	12
	L planum temporale	4.80	−58	−44	8
	LIFG	4.60	−44	10	24
	Left middle occipital gyrus	6.75	−44	−66	8
	Left inferior occipital gyrus	7.43	−24	−96	8
	Right middle occipital gyrus	8.90	50	−68	6
	Right inferior occipital gyrus	8.15	22	−94	−8
	Left thalamus	6.98	−18	−30	−4
	Right thalamus	6.11	20	−28	−2

The table shows a description of the comparison performed, a description of the region activated, the *T*-value of the maximally activated voxel and the corresponding MNI coordinates. Results are correct for multiple comparisons at $p<0.05$.

inspection at a lower, uncorrected threshold ($p<0.005$ uncorrected) showed increased activation in LIFG, but not in pSTS/MTG, in agreement with the ROI analysis. Also no activation was observed in pSTS/MTG at an even more liberal threshold of $p<0.01$ uncorrected.

Effective connectivity analysis

The PPI analysis with the time course of the a priori defined ROI in LIFG showed that effective connectivity from this region is increased in the Pant-Mism condition as compared to the Pant-Match condition with left pSTS, bilateral lateral occipital sulci, left cuneus, right calcarine sulcus and right inferior occipital sulcus (Fig. 5 and Table 6). The area in left pSTS overlaps with the cluster in this area that was found to be activated in the congruency contrast (Pant-mism>Pant-match) reported above (see Supplementary Fig. S1). We performed PPI analyses using the time course from this activation cluster in left pSTS. No connections with left inferior frontal cortex were present (also not at $p<0.01$ uncorrected), attesting to the unidirectionality of the effect (that is, from LIFG to pSTS). Neither did the cluster in pSTS/MTG that was activated in the whole-brain analysis show such an

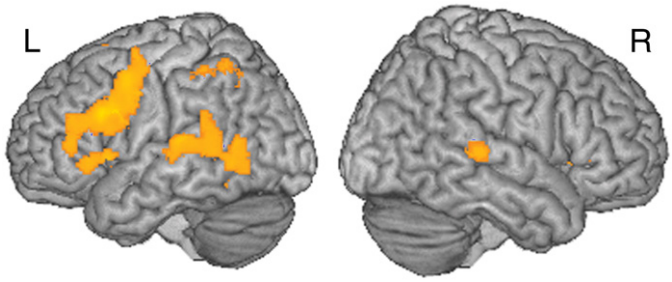

Fig. 4. Areas activated in whole-brain analysis to the Pant-Mism versus Pant-Match contrast. Map is thresholded at $p<0.05$, corrected for multiple comparisons, and overlaid on a rendered brain. This analysis generally confirms the ROI analysis with increased activation in left IFG and bilateral pSTS/MTG. The other activation clusters did not exhibit multimodal properties and are not discussed further (see main text). No areas were activated to the Gest-Mism versus Gest-Match comparison. However, at a lower statistical threshold ($p<0.005$ uncorrected), LIFG was also activated to the Gest-Mism versus Gest-Match contrast. This was not the case for pSTS/MTG (also not at $p<0.01$ uncorrected).

Table 5
Results of whole-brain analysis comparing Pant-Mism versus Pant-Match and Gest-Mism versus Gest-Match.

Region	T(max)	Coordinates (MNI)		
		x	*y*	*z*
Pant-Mism versus Pant-Match				
L posterior STS/MTG	4.72	−56	−46	6
	4.59	−56	−64	2
R posterior STS	5.94	62	−32	4
L inferior frontal gyrus/precentral sulcus	11.33	−40	10	22
L intraparietal sulcus	7.88	−34	−54	46
L insula	5.30	−42	24	−2
R insula	4.55	40	24	4
L and R cingulate sulcus	10.27	−8	10	58
		8	20	48
Gest-Mism vs. Gest-Match				
–	–	–	–	–

Displayed are an anatomical description of the region, the *T*-value of the maximally activated voxel in the region and the centre coordinates of the region in MNI space.

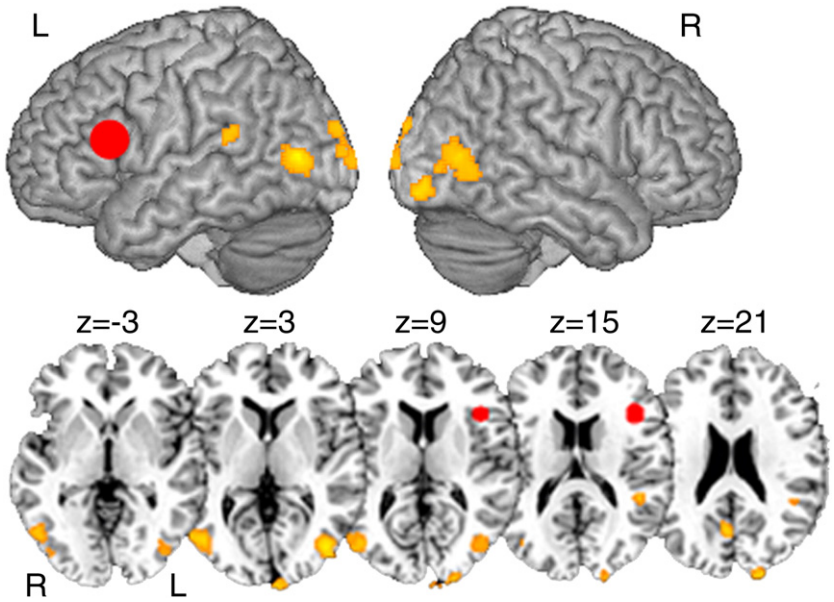

Fig. 5. Results of effective connectivity analysis taking the a priori defined region of interest in LIFG as seed region. The statistical map shows areas that are more strongly modulated by LIFG in the Pant-Mism condition as compared to the Pant-Match condition. This was the case for left pSTS, bilateral lateral occipital sulci, left cuneus, right inferior occipital sulcus and right calcarine sulcus. Map is thresholded at $p<0.05$, corrected for multiple comparisons and overlaid on a rendered brain. The rendered image is somewhat misleading since it displays activations at the surface of the cortex that are actually 'hidden' in sulci. Therefore, we also display the result on multiple coronal slices. In the latter view, localization of the activation in left pSTS is more straightforward. The cluster in pSTS overlaps with activation in this region in the whole-brain Pant-Mism>Pant-Match contrast (see Supplementary Fig. S1). No areas were found to be more strongly modulated by LIFG in the Gest-Mism as compared to Gest-Match condition (also not at $p<0.01$ uncorrected).

effect. Some of the other areas found activated in this analysis overlap with, or are in the vicinity of, previously reported 'Extrastriate Body Area' (Peelen et al., 2006). However, none of these latter areas showed multimodal response characteristics and we do not discuss them further. No areas showed effective connectivity with LIFG as a function of the Gest-Mism condition as compared to the Gest-Match condition. Neither were any areas found to be modulated at an uncorrected statistical threshold of $p<0.01$. A direct statistical comparison between effective connectivity from IFG in the Speech–Pantomime mismatch versus Speech–Pantomime match contrast as compared to the Speech–Gesture mismatch versus Speech–Gesture match contrast, showed a similar result.

The PPI analysis with the time course from the a priori defined ROIs in left and right pSTS/MTG, showed that connectivity from right pSTS was increased in the Pant-Mism condition as compared to Pant-Match condition in right inferior temporal sulcus and left superior occipital gyrus (Fig. 6 and Table 7). No areas showed effective connectivity with right pSTS/MTG as a function of Gest-Mism versus Gest-Match or with left pSTS/MTG as a function of the Gest-Mism versus Gest-Match or Pant-Mism versus Pant-Match. Informal inspection at uncorrected statistical thresholds ($p<0.01$ uncorrected) revealed that no effective connectivity was present from either of the pSTS/MTG ROIs onto LIFG.

Results summary

In summary, in the ROI analysis we found that all a priori defined regions of interest exhibited multimodal response characteristics (Beauchamp, 2005b). The congruency analysis showed that left and right pSTS/MTG were sensitive to congruence of pantomimes and speech, but not of gestures and speech, whereas LIFG was sensitive to congruence in both Speech–Pantomime and Speech–Gesture combinations. Testing the parametrically varying degree of congruence (as determined in a pretest) between Speech–Pantomime and Speech–Gesture combinations confirmed that these

Table 6
Results of effective connectivity analysis with time course from the a priori defined ROI in LIFG as seed region.

Contrast	Region	T(max)	Coordinates (MNI)		
			x	y	z
Pant-Mism vs. Pant-Match	L posterior STS	3.73	−46	−42	14
	L lateral occ. sulcus	8.63	−44	−73	9
	R lateral occ. sulcus	3.84	42	−71	14
	L cuneus	4.58	−8	−106	2
	R inf. occ. sulcus	6.87	30	−94	−14
	R calcarine sulcus	5.55	4	−66	20
Gest-Mism vs. Gest-match	–	–	–	–	–

An area in left pSTS overlapping with the area found in the main contrast in the whole-brain analysis was found to be modulated by LIFG, depending upon whether the condition was Pant-Mism versus Pant-Match (see Fig. S1 for a visualization of the overlap). No areas were influenced by LIFG depending upon whether the condition was Gest-Mism versus Gest-Match.

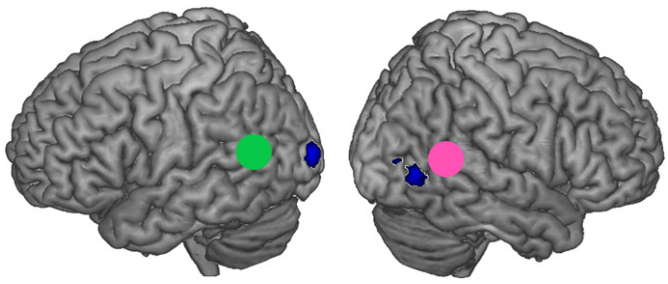

Fig. 6. Results of effective connectivity analysis taking the a priori defined regions of interest in left or right pSTS/MTG as seed region. The left hemisphere ROI is in green, the right hemisphere ROI is in violet. Map is thresholded at $p<0.05$, corrected for multiple comparisons and overlaid on a rendered brain. No areas were modulated by left pSTS/MTG in the Pant-Mism versus Pant-Match or Gest-Mism versus Gest-Match comparisons (also not at $p<0.01$ uncorrected). Right inferior temporal and left superior occipital gyrus were more strongly modulated by right pSTS/MTG in the Pant-Mism condition as compared to the Pant-Match condition (areas indicated in blue). No areas were found to be more strongly modulated by right pSTS/MTG in the Gest-Mism as compared to Gest-Match condition (also not at $p<0.01$ uncorrected).

Table 7
Results of effective connectivity analysis with time course from the a priori defined ROIs in left and right pSTS/MTG as seed region.

Contrast	Region	*T*(max)	Coordinates (MNI)		
			x	*y*	*z*
Left pSTS/MTG					
Pant-Mism vs. Pant-Match	–	–	–	–	–
Gest-Mism vs. Gest-match	–	–	–	–	–
Right pSTS/MTG					
Pant-Mism vs. Pant-Match	Right inferior temporal gyrus	7.03	42	−68	4
	Left superior occipital gyrus	6.90	−26	−96	6
Gest-Mism vs. Gest-match	–	–	–	–	–

The table displays regions that were influenced by left or right pSTS/MTG, depending upon whether the condition was Pant-Mism versus Pant-Match. No areas were influenced by left or right pSTS/MTG depending upon whether the condition was Gest-Mism versus Gest-Match.

areas were also sensitive to the degree of congruence. This rules out the alternative explanation that we did not observe an effect of Speech–Gesture combinations in left pSTS/MTG due to the greater spread of congruence in these stimuli as compared to Speech–Pantomime combinations. In the whole-brain analysis this pattern of results was repeated. Finally, we found that LIFG has stronger effective connectivity with pSTS during Pant-Mism condition as compared to Pant-Match condition. Such an influence of IFG onto pSTS was not observed for the Gest-Mism condition as compared to the Gest-Match condition. Posterior STS/MTG showed stronger connectivity with left middle occipital gyrus, left cuneus and right superior frontal sulcus during Pant-Mism combinations as compared to Pant-Match combinations.

Discussion

In this study we investigated the functional roles of posterior superior temporal sulcus/middle temporal gyrus and left inferior frontal gyrus during multimodal integration. Two types of action-language combinations were investigated: speech combined with co-speech gestures and speech combined with pantomimes. Spoken language and co-speech gestures are strongly and intrinsically related to each other, in the sense that they are produced together and that gestures cannot be unambiguously recognized or understood when they are presented without speech (e.g. Riseborough, 1981; Feyereisen et al., 1988; Krauss et al., 1991; McNeill, 1992; Beattie and Shovelton, 2002; Goldin Meadow, 2003; Kita and Özyürek, 2003; Kendon, 2004). This is not the case for pantomimes, which are often not produced together with speech and are easily understood without speech (Goldin Meadow et al., 1996). We found that areas involved in multimodal integration are differentially influenced by this difference in semantic relationship between the two input streams.

Specifically, we found that pSTS/MTG is only sensitive to congruence of simultaneously presented speech and pantomimes, but not to simultaneously presented speech and co-speech gestures. On the contrary, LIFG was modulated by congruence of both Speech–Gesture as well as Speech–Pantomime combinations. Below we discuss what these findings reveal about the functional roles of pSTS/MTG and LIFG in multimodal integration.

Posterior STS/MTG has been implicated in multimodal integration in a multitude of studies, for instance in integration of phonemes and lip movements (e.g. Calvert et al., 2000; Calvert, 2001; Callan et al., 2003, 2004; Skipper et al., 2007b), phonemes and written letters (van Atteveldt et al., 2004, 2007), objects and their related sounds (Beauchamp et al., 2004b; Taylor et al., 2006) and pictures of animals and their sounds (Hein et al., 2007; Hein and Knight, 2008). Here we show that this area is also involved in integration of information from meaningful actions (pantomimes) and verbs that describe the pantomime.

Also IFG has been found to be involved in semantic multimodal integration. For instance, this region is sensitive to semantic incongruity of the simultaneously presented picture of an animal and the sound of another animal (Hein et al., 2007) and to integration of non-existing objects ('fribbles') with sounds (Hein et al., 2007; Naumer et al., 2008). Moreover, in a large amount of language studies, LIFG has been repeatedly found to be involved in semantic processing in a sentence context (e.g. Friederici et al., 2003; Kuperberg et al., 2003; Hagoort et al., 2004; Rodd et al., 2005; Ruschemeyer et al., 2005; Davis et al., 2007; Hagoort et al., in press). This is also true when integrating extra-linguistic information such as gestures or pictures in relation to a previous sentence context (Hagoort and van Berkum 2007; Willems and Hagoort 2007; Willems et al., 2007; Straube et al., 2009; Tesink et al., in press; Willems et al., 2008a, 2008b).

What partially distinct roles do pSTS/MTG and LIFG play in multimodal integration? Neuroimaging literature suggests that pSTS/MTG plays its role in multimodal integration by mapping the content of two input streams onto a common object representation in long-term memory (Beauchamp et al., 2004b; Amedi et al., 2005; Beauchamp, 2005a). This explains why we find modulation of pSTS/MTG for speech and pantomimes and not for speech and co-speech gestures. The content of both the verbs and the pantomimes can be mapped onto a relatively stable, common conceptual representation of that action/word in memory. This is crucially not the case for co-speech gestures. The dependency of gestures on accompanying language necessitates that semantic integration happens only at a higher level of semantic processing than for input streams that can be mapped onto a representation lower in the cortical hierarchy. That is, integrating gestures with speech invokes the construction of a *novel* representation instead of mapping input streams onto an already existing representation. Our findings show that LIFG and not pSTS/MTG is involved in such higher level integration.

Converging evidence for this comes from two recent studies. First, it was found that IFG (but not pSTS/MTG) was involved in integration of novel associations of non-existing objects and sounds (Hein et al., 2007). On the contrary, both LIFG and pSTS/MTG were involved in integration of animal pictures and their sounds (Hein et al., 2007). Second, Naumer et al. found an interesting shift in the activation pattern related to training of bimodal object presentations. They scanned participants who observed non-existing objects ('fribbles') paired with artificial sounds, before and after training of sound-object pairings. Interestingly, in the pre-training data, bilateral IFG, but not pSTS/MTG, was found to be involved in multimodal integration. After training, both IFG as well as pSTS were found activated to bimodal presentation of the stimuli (as compared to the maximum of unimodal presentation). This is nicely in agreement with the suggestion from the present data that pSTS/MTG is involved in integration of bimodal stimuli for which a relatively stable pairing exists, but not when integration involves the creation of a novel pairing between the bimodal input streams.

Our effective connectivity results further illuminate the interplay between LIFG and pSTS/MTG during multimodal integration. That is, in reaction to a mismatching Speech–Pantomime combination, LIFG modulates activation levels in areas lower in the cortical hierarchy, most notably pSTS and an area in the vicinity of previously reported Extrastriate Body Area (EBA) (Peelen et al., 2006). This modulatory function of IFG has been suggested before (Skipper et al., 2007a; see Gazzaley and D'Esposito 2007 for overview) and is in line with the proposed function of this area in regulatory functions such as semantic selection/control/unification (Thompson-Schill et al., 1997; Badre et al., 2005; Hagoort 2005b,a; Thompson-Schill et al., 2005). In this scenario, LIFG and pSTS work together to integrate multimodal information, with a modulatory role of LIFG and a more integrative role for pSTS (in the sense of mapping the input streams onto a relatively stable common representation). This fits with the

finding that during multimodal integration, pSTS/MTG precedes activation in LIFG in time (Fuhrmann Alpert et al., 2008). Our findings show that LIFG can subsequently modulate pSTS. On the contrary, when integration does not involve pSTS, as was the case in the Speech–Gesture combinations, there is no such modulatory signal from LIFG to pSTS.

It might be misleading to draw a sharp distinction between modulation on the one hand and integration on the other hand. Hagoort (2005b,a) has characterized IFG's function as *unification*, which crucially implies both modulation of areas lower in the cortical hierarchy as well as integration of information. For instance, during sentence comprehension this area can maintain activation of conceptual representations for the sake of unification, as well as integrate incoming information into a wider, previous sentence or discourse context (see Hagoort 2005a; Hagoort et al., in press for discussion). Our present findings seem to be in line with such an account, in the sense that LIFG exhibits both modulatory as well as integrative functions, crucially depending upon the semantic relationship between the input streams. It is important to stress that the integrative function of LIFG involves constructing a novel representation, based upon the two input streams. As such, and as we argued above, integration processes in pSTS/MTG and LIFG are of a different nature.

An interesting difference between this and some other multimodal studies is that in our study, in pSTS/MTG, activation levels increased in response to mismatching stimulus combinations (see also Hein et al., 2007; Hocking and Price, 2008). In contrast, some multimodal integration studies report activation increases to *matching* stimulus combinations (Beauchamp et al., 2004b; van Atteveldt et al., 2004). Our pattern of results is in the opposite direction, but is commonly reported in studies that modulate the semantic integration load of a word into a preceding sentence context (e.g. Bookheimer, 2002; Friederici et al., 2003; Kuperberg et al., 2003; Hagoort et al., 2004; Rodd et al., 2005; Ruschemeyer et al., 2005; Davis et al., 2007; Willems et al., 2007, 2008b). An intriguing but speculative explanation is that the presence of language stimuli at and beyond the word level creates this difference. Future research should investigate this in a more systematic way.

A possible criticism to our study could be the use of a mismatch paradigm. The mismatch paradigm is widely used in the neurocognition of language and has been shown to successfully increase integration load of an item into a previous context (see Kutas and Van Petten, 1994; Brown et al., 2000 for review). Importantly, ERP studies show that the N400 effect is elicited by semantic anomalies as well as by more subtle semantic manipulations that do not invoke an anomaly (Kutas and Hillyard, 1984; Hagoort and Brown, 1994). Similarly, there are fMRI studies which find that similar neural networks show increased activation levels in paradigms which manipulate semantic integration load without using a mismatch paradigm (Rodd et al., 2005; Davis et al., 2007). In short, semantic anomalies are the end point of a continuum that embodies increased semantic processing load. Also studies of multimodal integration have successfully employed a mismatch paradigm (Beauchamp et al., 2004b; Hein et al., 2007; van Atteveldt et al., 2007; Fuhrmann Alpert et al., 2008; Hocking and Price, 2008). Moreover, all ROIs in our study were also activated above baseline during presentation of the *matching* Speech–Gesture and Speech–Pantomime combinations (Fig. 2; Table 2).

When we compared bimodal versus the combination of unimodal conditions (bimodal>unimodal-audio + unimodal-video) we also observed activation increases in (primary) auditory and visual cortices. A similar finding of auditory cortex was observed comparing speech combined with beat gestures to unimodal presentation of speech and beat gestures (Hubbard et al., 2009), as well as when comparing sound–picture pairings to unimodal presentations (Hein et al., 2007). Such effect was observed in visual cortices by Belardinelli et al. (2004) in response to a similar comparison. In contrast, van Atteveldt et al. (2004) observed congruency effects in auditory cortex, which was not replicated here. So it seems that bimodal presentation of stimuli leads to stronger activations in primary and non-primary auditory and visual cortex as compared to the combination of unimodal presentations. However, these areas are not sensitive to the semantic congruency in both processing streams. Hence, we refrain from implicating them in semantic integration.

In summary, we have shown that areas known to be involved in multimodal integration are also involved in integration of language and action information. Importantly, the relationship between language and action information crucially changes the areas involved in integration of the two information types.

Acknowledgments

Supported by a grant from the Netherlands Organization for Scientific Research (NWO), 051.02.040 and by the European Union Joint-Action Science and Technology Project (IST-FP6-003747). We thank Cathelijne Tesink and Nina Davids for help in creation of the stimuli and Caroline Ott for help at various stages of the project. Paul Gaalman is acknowledged for his expert assistance during the scanning sessions.

Appendix A

Transcription of speech segments and descriptions of pantomimes and gestures. Speech segments indicate the verbs (used in the Speech–Pantomime combinations) and the speech phrases (used in the Speech–Gesture combinations) that were used in the experiment. Under each Dutch speech description there is a translation in English. The brackets in Speech–Gesture pairs indicate where the stroke (the meaningful unit of the movement) (McNeill, 1992) of each gesture occurred in relation to the speech. Co-speech gestures were segmented and described according to conventions in McNeill (1992) as well as in Kita et al. (1998). In pantomime descriptions we also took McNeill's (1992) co-speech gesture description conventions as a guide.

Speech–Pantomime pairs

Speech (originals in Dutch (italics) with English translations)	Pantomime description
Typen	Selected fingers on both hands move up and
To type	down in a typing manner, palms facing down
Schudden	C hand shape, palm facing sideways, move up and
To shake	down
Schrijven	Index finger grasping thumb tip ('money'
To write	handshape), palm facing down moves laterally in small arcs
Scheuren	Both hands in 'money' handshape facing down
To tear	move away from each other on sagittal axis
Roeren	'Money' handshape pointing down moves in
To stir	circles
Kloppen	Fist hand moves back and forth away from the
To knock	body
iets opendraaien	Right hand claw shaped, palm facing down,
To unscrew	moves sideways repetitively over the left hand, while C shaped left hand palm facing sideways rests in place below the right hand
iets intoetsen	Left B hand facing towards body remains in place
To type in	and right index finger taps on the left hand
iets inschenken	Fist hand facing sideways moves up and then
To pour	down in an arc motion
Grijpen	One hand moves laterally from C handshape palm
To grasp	facing to the side to a closed fist
Gewichtheffen	Fist hand facing body moves up and down from
To lift weight	the elbow
Breken	Fist hands make a break motion from middle to
To break	the sides and down

Speech–Gesture pairs

Speech (originals in Dutch with English translations)	Gesture descriptions
en dan [komt 'ie aanlopen] And then he walks in	C shaped hand, palm oriented down moves laterally
dan [loopt 'ie snel weg] Then he quickly walks away	Inverted V handshape moves laterally
en [valt 'ie weer terug naar beneden] And … he falls down again	B shaped flat hand pointing away from body moves down vertically
hij zwaait [tegen de muur aan] He swings into the wall	B shaped flat hand moves horizontally making a downward arc
eh die komt eh [binnenlopen] Uh he uh comes and walks in	Both hands with index finger extended move forward away form body depicting walking manner
[is hij eh heel druk aan het schrijven en aan het rekenen] He is uh very busy writing and calculating	Both hands in fist handshape, palms facing down move back and forth
[dan gaan ze elkaar achterna zitten] Then they go and chase each other	Index finger pointing down makes a couple of circles
[loopt onder aan de regenpijp op en neer] Walks from one side to the other	Inverted V handshape moves laterally to the sides back and forth
[en die gaat naar beneden] And he goes down	Index finger moves vertically pointing down
en die rolt er[zo naar binnen] And he rolls in	Index and middle fingers extended move straight sagitally away from body
en de [ene die]$_a$ [smijt 'ie weg]$_b$ And the one he throws away	a. Index and middle fingers extended point to a location in front of the speaker b. B shaped flat hand moves horizontally in a sweeping manner
hij staat er vrolijk [aan te draaien] He is happily turning it	Fist shaped hand turns around a couple of times

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2009.05.066.

References

Amedi, A., von Kriegstein, K., van Atteveldt, N.M., Beauchamp, M.S., Naumer, M.J., 2005. Functional imaging of human crossmodal identification and object recognition. Exp. Brain Res. 166 (3–4), 559–571.

Badre, D., Poldrack, R.A., Pare-Blagoev, E.J., Insler, R.Z., Wagner, A.D., 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. Neuron 47 (6), 907–918.

Beattie, G., Shovelton, H., 2002. An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. British J. Psychol. 93 (2), 179–192.

Beauchamp, M.S., 2005a. See me, hear me, touch me: multisensory integration in lateral occipital–temporal cortex. Curr. Opin. Neurobiol. 15 (2), 145–153.

Beauchamp, M.S., 2005b. Statistical criteria in FMRI studies of multisensory integration. Neuroinformatics 3 (2), 93–113.

Beauchamp, M.S., Argall, B.D., Bodurka, J., Duyn, J.H., Martin, A., 2004a. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat. Neurosci. 7 (11), 1190–1192.

Beauchamp, M.S., Lee, K.E., Argall, B.D., Martin, A., 2004b. Integration of auditory and visual information about objects in superior temporal sulcus. Neuron 41 (5), 809–823.

Belardinelli, M.O., Sestieri, C., Di Matteo, R., Delogu, F., Del Gratta, C., Ferretti, A., Caulo, M., Tartaro, A., Romani, G.L., 2004. Audio–visual crossmodal interactions in environmental perception: an fMRI investigation. Cogn. Processes 5, 167–174.

Bookheimer, S., 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annu. Rev. Neurosci. 25, 151–188.

Brown, C.M., Hagoort, P., Kutas, M., 2000. Postlexical integration processes in language comprehension: evidence from brain-imaging research. The Cognitive Neurosciences. M. S. Gazzaniga. InMIT Press, Cambridge, Mass, pp. 881–895.

Buchel, C., Holmes, A.P., Rees, G., Friston, K.J., 1998. Characterizing stimulus–response functions using nonlinear regressors in parametric fMRI experiments. NeuroImage 8 (2), 140–148.

Callan, D.E., Jones, J.A., Munhall, K., Callan, A.M., Kroos, C., Vatikiotis-Bateson, E., 2003. Neural processes underlying perceptual enhancement by visual speech gestures. NeuroReport 14 (17), 2213–2218.

Callan, D.E., Jones, J.A., Munhall, K., Kroos, C., Callan, A.M., Vatikiotis-Bateson, E., 2004. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. J. Cogn. Neurosci. 16 (5), 805–816.

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

Calvert, G.A., Thesen, T., 2004. Multisensory integration: methodological approaches and emerging principles in the human brain. J. Physiol. Paris 98 (1–3), 191–205.

Calvert, G.A., Campbell, R., Brammer, M.J., 2000. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr. Biol. 10 (11), 649–657.

Clark, H.H., Gerrig, R.J., 1990. Quotations as demonstrations. Language 66, 764–805.

Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8 (2–3), 109–114.

Davis, M.H., Coleman, M.R., Absalom, A.R., Rodd, J.M., Johnsrude, I.S., Matta, B.F., Owen, A.M., Menon, D.K., 2007. Dissociating speech perception and comprehension at reduced levels of awareness. Proc. Natl. Acad. Sci. U. S. A. 104 (41), 16032–16037.

Duvernoy, H.M., 1999. The Human Brain: Surface, Three-dimensional Sectional Anatomy with MRI, and Blood Supply. Springer, Vienna.

Feyereisen, P., Van de Wiele, M., Dubois, F., 1988. The meaning of gestures: what can be understood without speech? Cahiers de Psychologie Cognitive/Curr. Psychol. Cogn. 8 (1), 3–25.

Friederici, A.D., Ruschemeyer, S.A., Hahne, A., Fiebach, C.J., 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cereb. Cortex 13 (2), 170–177.

Friston, K., 2002. Beyond phrenology: what can neuroimaging tell us about distributed circuitry? Annu. Rev. Neurosci. 25, 221–250.

Friston, K.J., Holmes, A., Poline, J.B., Price, C.J., Frith, C.D., 1996. Detecting activations in PET and fMRI: levels of inference and power. NeuroImage 4 (3 Pt 1), 223–235.

Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6 (3), 218–229.

Fuhrmann Alpert, G., Hein, G., Tsai, N., Naumer, M.J., Knight, R.T., 2008. Temporal characteristics of audiovisual information processing. J. Neurosci. 28 (20), 5344–5349.

Gazzaley, A., D'Esposito, M., 2007. Unifying prefrontal cortex function: executive control, neural networks and top-down modulation. In: Cummings, J., Miller, B. (Eds.), The Human Frontal Lobes. InGuildford, New York.

Gitelman, D.R., Penny, W.D., Ashburner, J., Friston, K.J., 2003. Modeling regional and psychophysiologic interactions in fMRI: the importance of hemodynamic deconvolution. NeuroImage 19 (1), 200–207.

Goldin Meadow, S., 2003. Hearing Gesture: How Our Hands Help Us Think. Belknap Press of Harvard University Press, Cambridge, MA, US.

Goldin Meadow, S., McNeill, D., Singleton, J., 1996. Silence is liberating: removing the handcuffs on grammatical expression in the manual modality. Psychol. Rev. 103 (1), 34–55.

Hagoort, P., 2005a. Broca's complex as the unification space for language. In: Cutler, A. (Ed.), Twenty First Century Psycholinguistics: Four cornerstones. InLawrence Erlbaum Associates Publishers, Mahwah, NJ,, pp. 157–172.

Hagoort, P., 2005b. On Broca, brain, and binding: a new framework. Trends Cogn. Sci. 9 (9), 416–423.

Hagoort, P., Brown, C., 1994. Brain responses to lexical ambiguity resolution and parsing. In: Frazier, L., Clifton Charles, J., Rayner, K. (Eds.), Perspectives in Sentence Processing. InLawrence Erlbaum Associates, Hillsdale, NJ, England, pp. 45–80.

Hagoort, P., van Berkum, J., 2007. Beyond the sentence given. Philos. Trans. R. Soc. Lond. B Biol. Sci. 362 (1481), 801–811.

Hagoort, P., Baggio, G., Willems, R.M., in press. Semantic unification. The Cognitive Neurosciences IV. M. S. Gazzaniga, MIT press: Cambridge, MA.

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M., 2004. Integration of word meaning and world knowledge in language comprehension. Science 304 (5669), 438–441.

Hein, G., Knight, R.T., 2008. Superior temporal sulcus—it's my area: or is it? J. Cogn. Neurosci. 20 (12), 2125–2136.

Hein, G., Doehrmann, O., Muller, N.G., Kaiser, J., Muckli, L., Naumer, M.J., 2007. Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. J. Neurosci. 27 (30), 7881–7887.

Hocking, J., Price, C.J., 2008. The role of the posterior superior temporal sulcus in audiovisual processing. Cerebral Cortex.

Holle, H., Gunter, T.C., Ruschemeyer, S.A., Hennenlotter, A., Iacoboni, M., 2008. Neural correlates of the processing of co-speech gestures. NeuroImage 39 (4), 2010–2024.

Hubbard, A.L., Wilson, S.M., Callan, D.E., Dapretto, M., 2009. Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. Hum. Brain Mapp. 30 (3), 1028–1037.

Kendon, A., 2004. Gesture: Visible Action as Utterance. Cambridge University Press, Cambridge.

Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., Konrad, K., Green, A., 2009. Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures. Neuropsychologia 47 (1), 169–179.

Kita, S., Özyürek, A., 2003. What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. J. Mem. Lang. 48 (1), 16–32.

Kita, S., van Gijn, I., van der Hulst, H., 1998. Movement phases in signs and co-speech gestures and their transcription by human coders. In: Wachsmuth, I., Frohlich, M. (Eds.), Gesture and Sign Language in Human–Computer Interaction. InSpringer-Verlag, Berlin, pp. 23–35.

Krauss, R.M., Morrel Samuels, P., Colasante, C., 1991. Do conversational hand gestures communicate? J. Pers. Soc. Psychol. 61 (5), 743–754.

Kuperberg, G.R., Holcomb, P.J., Sitnikova, T., Greve, D., Dale, A.M., Caplan, D., 2003. Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. J. Cogn. Neurosci. 15 (2), 272–293.

Kutas, M., Hillyard, S.A., 1984. Brain potentials during reading reflect word expectancy and semantic association. Nature 307 (5947), 161–163.

Kutas, M., Van Petten, C.K., 1994. Psycholinguistics electrified: event-related brain potential investigations. In: Gernsbacher, M.A. (Ed.), Handbook of Psycholinguistics. InAcademic Press, San Diego, CA, pp. 83–143.
Levene, H., 1960. Robust tests for equality of variances. In: Olkin, I., Ghurye, S.G., Hoeffding, W., Madow, W.G., Mann, H.B. (Eds.), Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. InStanford University Press, Palo Alto, CA, pp. 278–292.
McNeill, D., 1992. Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press, Chicago, IL, US.
McNeill, D., 2000. Language and gesture. Cambridge University Press, Cambridge.
McNeill, D., 2005. Gesture and Thought. University of Chicago Press, Chicago, IL, US.
McNeill, D., Cassell, J., McCullough, K.E., 1994. Communicative effects of speech-mismatched gestures. Res. Lang. Soc. Interact. 27 (3), 223–237.
Miller, E.K., 2000. The prefrontal cortex and cognitive control. Nat. Rev. Neurosci. 1 (1), 59–65.
Naumer, M.J., Doehrmann, O., Muller, N.G., Muckli, L., Kaiser, J., Hein, G., 2008. Cortical plasticity of audio–visual object representations. Cereb. Cortex.
Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B., 2005. Valid conjunction inference with the minimum statistic. NeuroImage 25 (3), 653–660.
Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113.
Özyürek, A., 2002. Do speakers design their cospeech gestures for their addressees?: The effects of addressee location on representational gestures. J. Mem. Lang. 46 (4), 688–704.
Özyürek, A., Willems, R.M., Kita, S., Hagoort, P., 2007. On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. J. Cogn. Neurosci. 19 (4), 605–616.
Peelen, M.V., Wiggett, A.J., Downing, P.E., 2006. Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. Neuron 49 (6), 815–822.
Riseborough, M.G., 1981. Physiographic gestures as decoding facilitators: three experiments exploring a neglected facet of communication. J. Nonverbal Behav. 5 (3), 172–183.
Rodd, J.M., Davis, M.H., Johnsrude, I.S., 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. Cereb. Cortex 15 (8), 1261–1269.
Ruschemeyer, S.A., Fiebach, C.J., Kempe, V., Friederici, A.D., 2005. Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. Hum. Brain Mapp. 25 (2), 266–286.
Skipper, J.I., Goldin Meadow, S., Nusbaum, H.C., Small, S.L., 2007a. Speech associated gestures, Broca's area and the human mirror system. Brain Lang. 101, 260–277.
Skipper, J.I., van Wassenhove, V., Nusbaum, H.C., Small, S.L., 2007b. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audio-visual speech perception. Cereb. Cortex 17 (10), 2387–2399.
Stein, B., Calvert, G., Spence, C., 2004. The Handbook of Multisensory Processes. MIT Press, Cambridge, MA.
Straube, B., Green, A., Weis, S., Chatterjee, A., Kircher, T., 2009. Memory effects of speech and gesture binding: cortical and hippocampal activation in relation to subsequent memory performance. J. Cogn. Neurosci. 21 (4), 821–836.
Taylor, K.I., Moss, H.E., Stamatakis, E.A., Tyler, L.K., 2006. Binding crossmodal object features in perirhinal cortex. Proc. Natl. Acad. Sci. U. S. A. 103 (21), 8239–8244.
Tesink, C.M., Petersson, K.M., Van Berkum, J.J., van den Brink, D., Buitelaar, J.K., Hagoort, P., in press. Unification of speaker and meaning in language comprehension: an fMRI study. J. Cogn. Neurosci.
Thompson-Schill, S.L., D'Esposito, M., Aguirre, G.K., Farah, M.J., 1997. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. Proc. Natl. Acad. Sci. U. S. A. 94 (26), 14792–14797.
Thompson-Schill, S.L., Bedny, M., Goldberg, R.F., 2005. The frontal lobes and the regulation of mental activity. Curr. Opin. Neurobiol. 15 (2), 219–224.
van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. Neuron 43 (2), 271–282.
van Atteveldt, N.M., Formisano, E., Blomert, L., Goebel, R., 2007. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb. Cortex 17 (4), 962–974.
Vigneau, M., Beaucousin, V., Herve, P.Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., Tzourio-Mazoyer, N., 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. NeuroImage 30 (4), 1414–1432.
Willems, R.M., Hagoort, P., 2007. Neural evidence for the interplay between language, gesture, and action: a review. Brain Lang. 101 (3), 278–289.
Willems, R.M., Özyürek, A., Hagoort, P., 2007. When language meets action: the neural integration of gesture and speech. Cereb. Cortex 17 (10), 2322–2333.
Willems, R.M., Oostenveld, R., Hagoort, P., 2008a. Early decreases in alpha and gamma band power distinguish linguistic from visual information during sentence comprehension. Brain Res. 1219, 78–90.
Willems, R.M., Özyürek, A., Hagoort, P., 2008b. Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. J. Cogn. Neurosci. 20 (7), 1235–1249.