

When Language Meets Action: The Neural Integration of Gesture and Speech

Roel M. Willems¹, Aslı Özyürek^{1,2} and Peter Hagoort^{1,3}

¹F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands, ²Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands and ³Nijmegen Institute for Cognition and Information, Radboud University Nijmegen, PO Box 9104, 6500 HE Nijmegen, The Netherlands

Although generally studied in isolation, language and action often co-occur in everyday life. Here we investigated one particular form of simultaneous language and action, namely speech and gestures that speakers use in everyday communication. In a functional magnetic resonance imaging study, we identified the neural networks involved in the integration of semantic information from speech and gestures. Verbal and/or gestural content could be integrated easily or less easily with the content of the preceding part of speech. Premotor areas involved in action observation (Brodmann area [BA] 6) were found to be specifically modulated by action information “mismatching” to a language context. Importantly, an increase in integration load of both verbal and gestural information into prior speech context activated Broca’s area and adjacent cortex (BA 45/47). A classical language area, Broca’s area, is not only recruited for language-internal processing but also when action observation is integrated with speech. These findings provide direct evidence that action and language processing share a high-level neural integration system.

Keywords: Broca’s area, fMRI, gestures, inferior frontal cortex, semantic processing

Introduction

Language and action are 2 core systems of human cognition. Moreover, they are often used together, as in pointing toward an object while producing its name. Despite this common co-occurrence, language and action are usually studied and conceived as separate domains within cognitive neuroscience. Consequently, very little is known about the neural circuitry underlying the integration of meaning from simultaneously perceived speech and action. Nevertheless, recent findings on the neurocognition of language semantics on the one hand (e.g., Pulvermuller 2005), and human action observation systems on the other (Decety et al. 1997; Rizzolatti and Arbib 1998; Rizzolatti et al. 2001; Molnar-Szakacs et al. 2005), suggest that the 2 systems recruit partly overlapping neural networks. In this study, we investigate the commonalities between language comprehension and action observation directly by presenting action and language-related stimuli simultaneously. To do so, we focus on one particular form of action that often co-occurs with language, namely cospeech gestures.

When someone talks to us, we not only hear speech but also see the speaker’s hand, mouth and body movements. In conversational settings, the brain therefore continuously integrates several streams of language and action-related information that contribute to the listener’s understanding of a speaker’s message. Among those sources of information, cospeech gestures constitute a particular form of action. That is, they have communicative content and are naturally produced

together with speech, contrary to, for instance, goal-directed object manipulations. As such, they are a prime example of actions that are recruited in the context of another domain of cognition (i.e., language). The present functional magnetic resonance imaging (fMRI) study investigates the neural locus of the integration of speech and action semantics as they co-occur simultaneously.

Previous studies investigating multimodal integration during communication have mostly focused on the relationship between lip movements and speech (Calvert 2001). Although both gestures and lip movements are examples of the natural co-occurrence of auditory and visual information during communication, they are fundamentally different with respect to their relationship to the speech they accompany. Whereas speech sounds and lip movements match with respect to form properties of language, there is no form matching between gestures and speech (McNeill 1992). Consider for example an upward hand movement in a climbing manner when a speaker says, “He climbed up the ladder.” Here, the gesture depicts the event as a whole, describing manner (“climb”) and direction (“up”) simultaneously. In speech, however, the message unfolds over time, broken up into smaller meaningful segments (i.e., the individual words *climb* and *up*). Because of these form differences (McNeill 1992), the mapping of speech and gesture information must occur at a higher, semantic level. Nevertheless, despite the fact that gestures express information in a different representational format than speech, the 2 modalities are systematically related in jointly conveying the speaker’s overall meaning (McNeill 1992; Clark 1996; McNeill 2000; Goldin Meadow 2003; Kita and Özyürek 2003; Kendon 2004). Thus, it has been claimed that speech and gesture are part of the same system of communication (McNeill 1992; Kendon 2004; Bernardis and Gentilucci 2006).

The systematic relationship between speech and gestures exists at 3 levels. First, there is semantic overlap between the representation in gestures and the meaning expressed in the concurrent speech, as in the “climb up” example above (e.g., McNeill 1992; Kita and Özyürek 2003). That is, speech and gesture usually convey similar or related information. Second, speech and gesture are temporally aligned to each other. A gesture phrase has 3 phases: the preparation, the stroke (semantically the most meaningful part of the gesture), and the retraction or hold (McNeill 1992). Studies have shown that the onset of the gesture (i.e., preparation) usually precedes the onset of the relevant speech segment by less than a second (Butterworth and Shovelton 1978; Morrel Samuels and Krauss 1992). More importantly, in most speech-gesture pairs the stroke coincides with the relevant speech segment (McNeill 1992). Finally, it has been shown that the spontaneous use of

gestures has a similar function as speech, namely to communicate the intended message to the addressee (e.g., Özyürek 2002; Kendon 2004; Melinger and Levelt 2004).

Furthermore, behavioral studies on speech and gesture comprehension have shown that listeners/viewers integrate information from gesture into their semantic interpretation of the speech input (Thompson and Massaro 1986; Thompson and Massaro 1994; Beattie and Shovelton 1999; Kelly et al. 1999). Listeners/viewers pick up information coming from gestures in naturally occurring situations when information is expressed only in gesture but not in the concurrent speech (Church and Goldin Meadow 1986; Kelly and Church 1998; Goldin Meadow and Momeni Sandhofer 1999; Goldin Meadow et al. 1999; Singer and Goldin Meadow 2005) and even in cases when gestures contradict the information simultaneously conveyed in speech (McNeill et al. 1994; Cassell et al. 1999).

Recently, few studies that have investigated brain responses with electrophysiological recordings (ERPs) during speech and gesture comprehension show that gestures evoke semantic processing. Kelly et al. (2004) found that ERPs to spoken words (targets) are modulated when the words are preceded by gestures (primes) containing information about the size and shape of objects that the target words referred to. Compared with words that matched the gesture primes, mismatching words evoked an early P1/N2 effect, followed by an N400 effect. On the basis of these findings Kelly et al. (2004) claimed that the gesture primes influenced word comprehension, first at the level of “sensory or phonological” processing and later at the level of semantic processing. In a study by Wu and Coulson (2005), it was found that incongruous gestures shown without speech and following cartoon images elicited a negative-going ERP effect around 450 ms compared with congruous gestures. In addition, it was observed that incongruous words following the cartoon-gesture pairs elicited an N400 effect. Holle and Gunter (2006) presented spoken sentences in which an ambiguous word was combined with a pantomimic gesture that cued one meaning of the ambiguous word. An N400 effect was found to a word later in the sentence when that word was incongruous to the meaning of the ambiguous word cued by the gesture. The authors concluded that a gesture cue can disambiguate the meaning of an ambiguous word.

With respect to neuroimaging, so far no fMRI studies have investigated the perception of cospeech gestures. However, many neuroimaging studies have investigated other types of action observation. These include the observation of pantomimes (Decety et al. 1997; Gallagher and Frith 2004), of hand emblems (Nakamura et al. 2004), of simple finger movements (Iacoboni et al. 1999; Koski et al. 2002; Molnar-Szakacs et al. 2005), and of actions toward objects (Hari et al. 1998; Nishitani and Hari 2000; Buccino et al. 2001; Grezes et al. 2003; Hamzei et al. 2003). Crucially, in all these studies actions were presented in isolation. Different neural responses in areas involved in action observation have been reported in different task settings (e.g., “passive observation” vs. “observe to imitate”). However, it is unknown to what extent these areas can be modulated by a language context. Nevertheless, a direct link between the language and action domains has been proposed, mainly inspired by neural findings in the monkey. When a monkey observes an action, neurons in areas that are thought to be homologous to human language areas are activated (Rizzolatti and Arbib 1998; Arbib 2005). Because classical language areas (Broca’s area) are also found activated in human

action observation, some have speculated about gestural communication as an immediate precursor of language in evolution (Rizzolatti and Arbib 1998; Arbib 2005; Nishitani et al. 2005; cf. Aboitiz and Garcia 1997; Bosman et al. 2005; Aboitiz et al. 2006).

In contrast to the absence of fMRI studies on the processing of cospeech gestures, a number of fMRI studies on language processing beyond the single word level are available. The language studies that have examined the neural networks underlying the semantic integration of word meaning into a representation of the preceding part of the utterance mostly used a mismatch paradigm. In this paradigm, the semantic integration load of a word’s meaning in relation to the preceding speech context is manipulated. fMRI studies using this approach found that language-internal semantic violations result in stronger activation in left superior temporal and left inferior frontal areas compared with a matching (i.e., semantically correct) control condition (Ni et al. 2000; Kuperberg et al. 2003).

Finally, there have been recent neuroimaging studies investigating sign language comprehension. Even though sign languages also use actions for communicative expressions like cospeech gestures, they differ in important ways from cospeech gestures, based on the fact that signs are lexicalized and produced in hierarchic combinations (McNeill 1992; Goldin Meadow 2003). A few fMRI studies investigated sentence comprehension in deaf signers (Neville et al. 1998; MacSweeney et al. 2002; Newman et al. 2002; MacSweeney et al. 2004; MacSweeney et al. 2006). These have shown that processing sentences in sign language activates a network of inferior frontal and temporal areas, which strongly overlaps with areas involved in sentence comprehension in hearing nonsigning individuals (see Corina and Knapp 2006; Emmorey 2006 for review). However, despite using the same visuospatial domain of expression, it is unknown and unclear whether cospeech gestures will activate the same areas.

In order to bridge the gap between these separate lines of research in the action and language domains, we investigated if similar neural systems are involved when semantic information conveyed through action or language needs to be integrated into the preceding context. We addressed this question by investigating which brain regions are responsive to variations of the semantic relationship between a gesture and/or a spoken word and the preceding part of a spoken sentence. In our study we presented participants with spoken sentences in which we manipulated the semantic “fit” of a verb (language) and/or a gesture (action) to the preceding sentence context (Table 1, Fig. 1). As found in previous language studies (Kutas and Hillyard 1980; Kutas and Hillyard 1984; Hagoort and Brown 1994), semantic integration load was expected to vary with this manipulation, which is commonly employed in neuroimaging studies of language (e.g., Kuperberg et al. 2000, 2003; Ni et al. 2000; Friederici et al. 2003; Hagoort et al. 2004; Ruschemeyer et al. 2006). In this way, regions specific for speech and gesture processing, as well as areas common to the integration of both information types into the prior sentence context could be identified. If integrating semantic information from both gesture and language into a broader sentence context activates the same areas, this would be direct evidence that the 2 systems recruit overlapping neural networks.

We call the critical verb or gesture which is semantically less fitting the previous sentence context “mismatch.” It is important to note that the term “mismatch” is used here in a different

Table 1

An example of the stimulus materials

Example sentence in Dutch (Critical words (L+/L-) underlined):"De artikelen die hij op het boodschappenlijstje schreef/sloeg moest hij niet vergeten."**Correct English translation**"He should not forget the items that he wrote/hit on the shopping list."*Correct condition (literal English translation)*G+L+: The items that he on the shopping list wrote should he not forget.
(write)*Language "mismatch"*G+L-: The items that he on the shopping list hit should he not forget.
(write)*Gesture "mismatch"*G-L+: The items that he on the shopping list wrote should he not forget.
(hit)*Double "mismatch"*G-L-: The items that he on the shopping list hit should he not forget.
(hit)

Note: In brackets [] is a verbal description of the iconic gesture. Gestures were displayed time locked to the onset of the verb (underlined). All stimuli were in Dutch, a literal translation in English is provided. Note that due to a difference in word order the English translation is ungrammatical. Note that the condition coding (G+L+, G+L-, etc.) refers to the match/ "mismatch" of either the verb (language: L) or the gesture (gesture: G) to the part of the sentence preceding the verb that is underlined. "Mismatches" are indicated in bold.

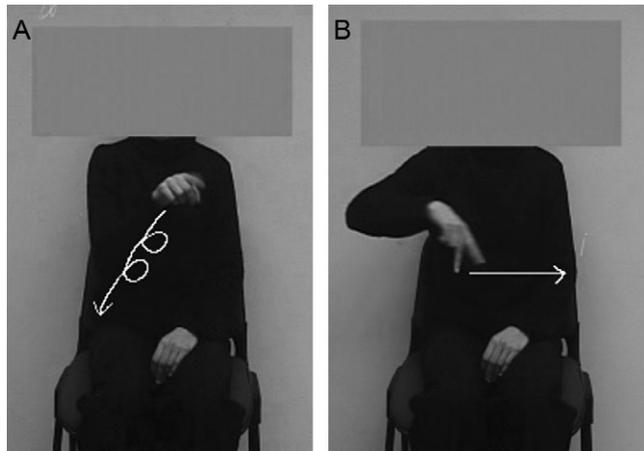


Figure 1. Two examples of the iconic gestures that were used. Depicted is one frame of the (A) "Roll down" gesture, and one frame of the (B) "Walk across" gesture. The line and arrow indicate the movement made by the hand. The original stimuli were in color.

sense than in other studies in the speech and gesture literature (e.g., Church and Goldin Meadow 1986), where it is called a mismatch when gesture conveys *additional*—not incongruent—information compared with speech.

The manipulation of the semantic fit in our materials resulted in 4 conditions (see Table 1, Fig. 1): correct condition, language "mismatch" condition, gesture "mismatch" condition, double "mismatch" condition. In the language "mismatch" the critical verb was harder to fit semantically to the preceding context, whereas the co-occurring gesture matched the sentence context. In the gesture "mismatch" condition the gesture was harder to integrate to previous context, whereas the critical verb matched the spoken sentence context. In the double "mismatch" condition both the gesture and the word were difficult to integrate to previous sentence context. Note that in the language and gesture "mismatch" conditions the critical verb and the overlapping gesture were locally incompatible (e.g., Speech: Write; Gesture: Hit, and vice versa), whereas in the double "mismatch" condition they were locally consistent (e.g.,

both Hit). Even though our study mainly targeted the effects of global sentence-level integration, this extra manipulation allowed us to check if our findings could be attributed to locally incompatible information of speech and gesture instead of to the context effects that we intended to study. The double "mismatch" condition should elicit similar effects as the language and gesture "mismatch" conditions, if what we are testing is really a global, sentence-level effect. Note that in our materials an increase of semantic integration load does not necessarily involve a strict semantic violation. That is, the critical word always fits the preceding sentence context less well in the "mismatch" conditions compared with the correct condition but is often not impossible as a continuation of the sentence. The condition label "mismatch" thus refers to cases where the continuation is pragmatically less plausible than in the correct condition but not necessarily impossible.

In particular, in this study we test the following specific hypotheses. The first concerns theories about the relation between speech and gesture systems. If speech and gesture are part of the same system of communication or interact at a high level of cognitive processing as is claimed on the basis of behavioral findings (McNeill 1992; McNeill 2000; Goldin Meadow 2003; Kita and Özyürek 2003; Kendon 2004), we expect the gesture and the language "mismatch" conditions to activate overlapping areas. If, however, cospeech gestures are considered not to have a communicative function (see Krauss et al. 1991), the "mismatching" gesture will not elicit similar effects as a "mismatch" in the language domain. Furthermore, neural overlap would provide further evidence for the claim that in the human brain there is a strong link between action and language systems (Rizzolatti and Arbib 1998; Nishitani et al. 2005).

Our second prediction concerns the fact that we expect the overlapping area of activation of language and gesture "mismatch" to include Broca's area and adjacent cortex. This is in relation to a recent proposal (Hagoort 2003; Hagoort et al. 2004; Hagoort 2005b), in which Broca's area and adjacent cortex (including Brodmann area, BA 47, 45, 44, and the ventral part of BA 6) in the left hemisphere serve as a unification space for language, with a focus in BA 45/47 for the unification of semantic information. During unification, lexical information retrieved from memory (i.e., from the mental lexicon) is integrated into a unified representation of a multiword utterance, such as a sentence. It is still an open question as to whether this unification space is specific for language or whether it integrates information across different domains of cognition. If Broca's area and adjacent cortex act as the general (not domain-specific) unification space for language and action, we predict left inferior frontal cortex to be activated stronger with higher semantic integration load of speech and gesture information. Specifically, based upon previous research we predict BA 45 and 47 to show increased activation with an increase in semantic integration load (Bookheimer 2002; Hagoort 2005b).

Third, we investigate whether and how regions of the human action recognition network are modulated by a language context. We focus on the neural processing of hand actions in premotor cortex (BA 6) and parietal cortex. Previous work has shown that part of the motor system "resonates" in a mirror-like fashion in response to the observation of actions (Rizzolatti et al. 2001; Nishitani et al. 2005). That is, the observation of an action triggers similar neural activity as executing an action. This

“neural simulation” of actions may underlie the understanding of actions performed by others (Jeannerod 2001; Rizzolatti et al. 2001; Nishitani et al. 2005). Studies in which actions were presented in isolation found modulations of the premotor cortex depending on the task, that is, whether participants observed actions with the intention to imitate versus passive observation of actions (Grezes et al. 1999; Molnar-Szakacs et al. 2005). Parietal regions are reported to be specifically modulated by object-related versus non object-related actions (Buccino et al. 2001), and by biologically impossible versus possible actions (Costantini et al. 2005). In this study, we seek to answer whether motor-related areas, besides being modulated by task setting and type of actions, can also be influenced by a language context. We hypothesize that part of the action recognition system will be more strongly activated when the semantic content of an action cannot be easily integrated into a broader context, that is, in the gesture “mismatch” condition. This would provide evidence that the action recognition system not only automatically codes features of observed actions but that it is also influenced by a previous semantic context provided by the language system.

Materials and Methods

Participants

Sixteen healthy volunteers ($N = 16$; 8 female; mean age = 24.1 years, range: 18–33) with normal or corrected to normal vision and normal hearing participated in the study. All participants were right handed (Oldfield 1971) and had Dutch as their native language. None of the participants had any known neurological impairment. Participants gave written informed consent in accordance with the declaration of Helsinki. The participants were paid for participation. The study was approved by the local ethics committee.

Stimulus Materials

The materials consisted of 640 items of spoken sentences that were accompanied by cospeech gestures. The sentences formed 160 sentence pairs. The members of a pair were identical up until the critical verb. Half of the sentences contained a critical verb that matched the preceding context. In the other half, the critical verb was semantically anomalous to the prior sentence context. Overall, 12 different critical verbs were used (supplementary information, available online). The sentences had an average duration of 3720 ms (standard deviation [SD] = 81), and the critical verbs had an average duration of 322 ms (SD = 85 ms). All sentences were spoken at a normal rate by a female speaker, recorded in a sound attenuated booth and stored onto disk.

The spoken sentences were combined with 12 iconic gestures (Fig. 1; supplementary information, available online). Iconic gestures are a class of gestures that speakers spontaneously use as they talk about spatial and activity related aspects of events (e.g., using wiggling fingers moving horizontally while talking about someone walking; McNeill 1992; Kita and Özyürek 2003). The iconic gestures used in this study were based on a larger database collected to investigate speakers’ natural and spontaneous use of speech and gestures in narratives of spatial events (Kita and Özyürek 2003). For the purposes of this study, 12 of these gestures were selected and modeled by one native female Dutch speaker with the requirement that they resembled spontaneous gestures in this database. The purpose behind using modeled gestures instead of natural ones was to be able to keep external factors constant across different gestures. In order to match the speed and length of the gesture phases (e.g., the stroke) as closely as possible to naturally occurring iconic gestures, we asked our model to produce concurrent sentences originally used in the narrative database as she was performing the gestures. During editing the audio was removed from the movie. Movies were edited using Adobe Premier (version 6.0; Adobe Systems Inc., San Jose, CA; <http://www.adobe.com>). The preparation and the retraction phase of each gesture were removed, leaving the stroke. Previous research has shown that especially the stroke conveys the

meaning of a gesture (McNeill 1992; McNeill 2000; Goldin Meadow 2003; Kita and Özyürek 2003; Kendon 2004). By isolating the gesture stroke phase, we eliminated differences among gestures that were due to the fact that for some gestures hand shape might reveal information before the stroke began, and that some gestures might have longer preparation time than others. The average length of the strokes was 767 ms (SD = 284 ms). Finally, the face of the model was blocked to eliminate the contribution of information coming from the lips.

The gestures corresponded to the meaning of the critical verbs. They were combined with the sentences in such a way that in half of the items the gesture matched the preceding sentence context, and in the other half it “mismatched” the preceding sentence context. This resulted in a total of 160 stimulus quartets (Table 1). In sum, there were 4 experimental conditions (Table 1): correct condition (gesture [G] +, language [L] +); language “mismatch” condition (G+L-); gesture “mismatch” condition (G-L+); double “mismatch” condition (G-L-).

The gesture movies and the sentence files were combined using the Adobe Premier (version 6.0) and After Effects software (version 5.5; Adobe Systems Inc., San Jose, CA <http://www.adobe.com>). For each movie file, the onset of the gesture stroke was temporally aligned with the onset of the critical verb because in 90% of natural speech-gesture pairs the stroke coincides with the relevant speech segment (McNeill 1992). For verbs with a separable prefix, the alignment point was not word-onset but the body of the verb following the prefix. The latter was the case for 44 sentences. Additional still frames with the hand resting on the lap were added to the part of the sentence before the critical verb, and the last frame of the stroke was elongated until the end of the sentence.

Four different stimulus lists were created, to distribute the 4 versions of each item equally over the 4 lists. This was done in such a way that all 4 lists contained an equal number of items per condition. Each list was presented to a quarter of the participants. As a result, none of the participants were presented with more than one version of the stimulus items, that is, every participant was presented with only one item from a quartet as in Table 1.

Experimental Design and Procedure

Forty items per condition were presented, resulting in a total of 160 items per participant. The items were presented in an event-related design, in a pseudorandomized order with the constraints that no more than 2 items of the same condition were presented after each other. The 4 stimulus lists were presented in normal or reversed order, resulting in 8 stimulus lists that were evenly distributed across male and female participants.

Stimuli were presented using the Nijmegen Experiment Setup software (NESU, MPI for Psycholinguistics; Nijmegen, The Netherlands <http://www.mpi.nl/world/tg/experiments/nesu.html>). The visual content of the movies was presented through an Eiki LC-X986 TFT-LCD projector outside the scanner room at a refresh rate of 60 Hz. Participants watched the screen via a nonmagnetic mirror mounted to the head coil. The movies subtended 10 cm (height) \times 11.8 cm (width) and were shown at a viewing distance of 80 cm. Speech was presented to the participants through nonmagnetic headphones (Commander XG, Resonance Technology Inc., Northridge, CA; <http://www.mrvideo.com>), which dampened scanner noise.

Participants were instructed to carefully listen to the sentences and watch the movies. They were told that they would receive questions about the items after the experiment. Before the beginning of a run, each participant received 2 practice runs consisting of 5 practice items each. These items were also used to adjust the volume level of the sentences. Therefore, the scanner was switched on during the practice items and participants were asked to indicate whether the volume should go up or down. The volume level that suited each participant best was used in the following experimental run. The functional data acquired during the practice runs were not used in the analysis.

fMRI Data Acquisition

fMRI was performed on a Siemens Magnetom Trio scanner (Siemens Medical Systems, Erlangen, Germany) with 3-Tesla magnetic field strength. Functional data were acquired with echo planar images in 32 transversal slices (repetition time [TR] = 2230 ms echo time [TE] = 30 ms; flip angle = 80°; slice thickness = 4 mm; field of view [FoV] = 224

mm, voxel resolution = 3.5 × 3.5 mm). Slices were positioned to cover the participant's whole brain. Intertrial interval was 2 or 3 scanner volumes (TRs), and the onset of each trial was synchronized to a scanner pulse. Sentence onset was effectively jittered by adding 0, 500, or 1000 ms (mean = 500 ms) to the trial onset (Josephs et al. 1997; Dale 1999; Miezin et al. 2000).

After the functional run, for each participant an anatomical scan was made using a high resolution T_1 -weighted 3D-MPRAGE sequence consisting of 192 sagittal slices (TR = 2300 ms; TE = 3.93 ms; FoV = 256 mm; slice thickness = 1 mm).

Data Analysis

Data were analyzed using Brainvoyager QX (Brain Innovation, Maastricht, The Netherlands; <http://www.brainvoyager.com>). The first 5 volumes of every functional run were discarded from the analysis to minimize T_1 -saturation effects. Preprocessing involved rigid body transformations of all volumes to the first volume to correct for small head movements, slice scan time correction, linear trend removal, and high-pass temporal filtering of 3 or fewer cycles per time course. The functional data of each run were coregistered to the anatomical data and were interpolated to a 1 × 1 × 1 mm voxel size. Subsequently, anatomical and functional data were transformed into stereotaxic space as defined by Talairach and Tournoux (1988). The functional data were spatially smoothed with a Gaussian filter kernel of 12-mm full-width half maximum (Xiong et al. 2000).

Regions of Interest Analyses

As described in the introduction, we had specific hypotheses for the anterior part of the left inferior frontal cortex (BA 45/47), the premotor cortex (BA 6), and the (inferior) parietal cortex. We, therefore, performed region of interest (ROI) analyses in these regions. A meta-analysis by Bookheimer (2002) showed that semantic processing is centered around [$x\ y\ z$] [-42 25 4] (Talairach and Tournoux 1988), with a mean distance of the local maxima to this center coordinate of 15 mm (Pettersson et al. 2004). Accordingly, a spherical ROI around [$x\ y\ z$] [-42 25 4] (Talairach and Tournoux 1988), with a radius of 15 mm was created. (In addition, we defined an ROI for left BA 45 based on an observer-independent cytoarchitectonic probability map; Amunts et al. 1999; Eickhoff et al. 2005. We included the voxels that had a probability of 50% or higher to fall within the borders of left BA 45. This region was converted from anatomical MNI space into stereotaxic Talairach space (Talairach and Tournoux 1988) by applying a nonlinear transformation [<http://www.mrc-cbu.cam.ac.uk/Imaging/Common/mnispace.shtml>]. A similar probability map for left BA 47 does not yet exist. By including this anatomically defined ROI for BA 45 we have an additional check on the validity of the ROI analysis that is solely based upon functional data from previous studies.) To avoid including air-tissue boundaries in our ROI, the inferior 3 mm of the sphere were not taken into the ROI.

The ROI for the premotor cortex (left and right BA 6) was defined on the basis of an observer-independent cytoarchitectonic probability map (Eickhoff et al. 2005), by including voxels that had a probability of 50% or higher to fall within the borders of BA 6. This region was converted from anatomical Montreal Neurological Institute (MNI) space into stereotaxic Talairach space (Talairach and Tournoux 1988) by applying a nonlinear transformation (<http://www.mrc-cbu.cam.ac.uk/Imaging/Common/mnispace.shtml>).

For the parietal cortex an ROI was constructed by averaging the local maxima of studies in which passive action observation was contrasted to a low level baseline (Grezes et al. 1999; Iacoboni et al. 1999; Buccino et al. 2001; Hamzei et al. 2003; Buccino et al. 2004; Costantini et al. 2005). This average was [$x\ y\ z$] [-35 -43 49] (Talairach and Tournoux 1988) for the left hemisphere and [$x\ y\ z$] [41 -38 52] (Talairach and Tournoux 1988) for the right hemisphere both in the vicinity of the intraparietal sulcus. The mean distances of the local maxima to these center coordinates were 16 mm (left hemisphere) and 9 mm (right hemisphere). Accordingly, 2 spherical ROIs were created around these averaged maxima with a radius of 16 and 9 mm, respectively (Supplementary Fig. S1 shows the ROIs in a glass brain).

Statistical analysis was done in the context of the General Linear Model. A model with the 4 experimental conditions (Table 1), in which events were modeled as the duration of the whole sentence convolved

with a canonical, 2 gamma hemodynamic response function (Friston et al. 1998) was tested in each participant's data separately. First, the average activation levels (beta weights) were estimated separately for each participant and condition in the a priori defined ROIs. Subsequently, paired t -tests ($df = 15$) to test for significant differences between conditions were applied to the estimated activation levels. Tested contrasts were language "mismatch" versus correct condition (G+L- vs. G+L+), gesture "mismatch" versus correct condition (G-L+ vs. G+L+), and double "mismatch" versus correct condition (G-L- vs. G+L+). All contrasts were tested two-sided.

Whole-Brain Analysis

In addition to testing condition effects in the ROIs, we also tested for the presence of other areas that were differentially activated by the experimental conditions. For this purpose, we performed a whole-brain random effects analysis, with the 4 conditions convolved with a canonical, 2 gamma hemodynamic response function (Friston et al. 1998) as our model. Individual contrast maps were taken to a second level analysis in which for each individual voxel the mean value of a contrast was tested against zero using the student's t -distribution with $df = 15 (n - 1)$. To control for the multiple comparisons problem introduced by the massive univariate approach taken, a voxel-wise intensity threshold ($P < 0.003$) was combined with a cluster extend threshold of $R > 41$ contiguous $3 \times 3 \times 3$ mm voxels, to control for false positives at an alpha level of $P < 0.05$ (Forman et al. 1995). All contrasts were tested two-sided.

Results

ROIs Analyses

First we explored if left inferior frontal cortex responds differently to action or language information that fits less easily into a sentence context than in the correct baseline condition. In the ROI in left inferior frontal cortex (BA 45/47) all tested contrasts revealed significant differences (Table 2, Fig. 2A): the language "mismatch" versus correct condition (G+L- vs. G+L+: $t(15) = 2.59, P < 0.02$), the gesture "mismatch" versus correct condition (G-L+ vs. G+L+: $t(15) = 2.53, P < 0.02$), and the double "mismatch" versus correct condition (G-L- vs. G+L+: $t(15) = 2.19, P < 0.04$). (The effects for the anatomically defined left BA 45 had a highly similar pattern. Again, significant differences were obtained between the language "mismatch" and correct condition (G+L- vs. G+L+: $t(15) = 3.24, P < 0.005$), the gesture "mismatch" and correct condition (G-L+ vs. G+L+: $t(15) = 3.65, P < 0.002$), and the double "mismatch" and correct condition (G-L- vs. G+L+: $t(15) = 3.11, P < 0.007$). See Supplementary Figure S2).

Second, the response of the motor system known to be involved in action observation was tested. In the ROI in premotor cortex (left and right BA 6), no significant differences were present between the language "mismatch" and the correct condition (G+L- vs. G+L+: $t(15) = 0.20, P < 0.84$), nor between the double "mismatch" and the correct condition (G-L- vs. G+L+: $t(15) = 0.79, P < 0.44$). In contrast, the gesture "mismatch" differed significantly from the correct condition (G-L+ vs. G+L+: $t(15) = 2.64, P < 0.02$; Table 2, Fig. 2B). In the left and right parietal ROIs, there was only a marginally significant effect in the left hemisphere for the contrast gesture "mismatch" versus correct condition (G-L+ vs. G+L+: $t(15) = 1.88, P < 0.08$; Table 2, Fig. 2C).

Whole-Brain Analysis

Subsequently, a more exploratory analysis was performed over the whole brain by testing for areas differentially activated by the contrasts of interest. (For general interest, we performed an analysis testing for areas significantly different from zero in the

Table 2

Activations in ROIs

Region	Centre coordinates			Contrast	<i>T</i>	df	<i>P</i> <
	<i>x</i>	<i>y</i>	<i>z</i>				
Left inferior frontal cortex (BA 45/47)	-42	25	9	Language mismatch versus correct	2.59	15	0.02
				Gesture mismatch versus correct	2.53	15	0.02
				Double mismatch versus correct	2.19	15	0.04
Premotor cortex (left and right BA 6)	0	-11	57	Language mismatch versus correct	0.20	15	0.84
				Gesture mismatch versus correct	2.64	15	0.02
				Double mismatch versus correct	0.79	15	0.44
Left parietal	-35	-43	49	Language mismatch versus correct	1.67	15	0.12
				Gesture mismatch versus correct	1.88	15	0.08
				Double mismatch versus correct	0.37	15	0.72
Right parietal	41	-38	52	Language mismatch versus correct	-1.10	15	0.29
				Gesture mismatch versus correct	1.50	15	0.16
				Double mismatch versus correct	-1.13	15	0.28

Note: The *T* values reflect differences between the averaged activation levels elicited by the various conditions. Regions were defined on the basis of previous functional imaging results (left inferior frontal cortex and left and right parietal regions) or on the basis of a cytoarchitectonic probability map (left and right BA 6). Centre coordinates are in stereotaxic space (Talairach and Tournoux 1988). Significant *T* and *P* values are in bold.

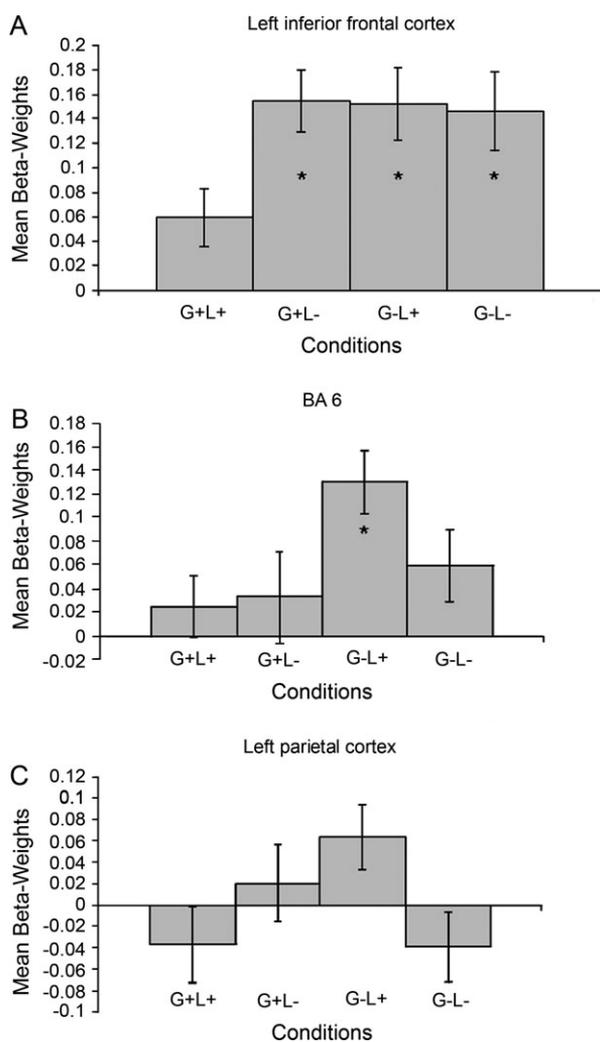


Figure 2. Gesture and speech in a sentence context. Mean activation levels (beta weights) for the 4 experimental conditions in (A) left inferior frontal cortex (BA 45/47), (B) left and right BA 6, and (C) left inferior parietal cortex. The activation levels are averaged over participants. An asterisk indicates a significant difference of the activation level of that condition compared with the correct condition (G+L+), at an alpha level of $P < 0.05$. See Table 2 for specific statistics. Error bars are standard error of the mean (SEM).

full model, reflecting areas responding to all conditions. The result of this analysis is provided in Supplementary Figure S3). In this whole-brain analysis, the comparison between the language “mismatch” condition and the correct condition resulted in significant activations in the left inferior frontal sulcus extending into the precentral sulcus, in the posterior part of the left superior temporal sulcus, and in the superior part of the left intraparietal sulcus (Table 3, Fig. 3A). For the gesture “mismatch” condition compared with the correct condition, we found significant activations in the left inferior frontal sulcus and in 2 areas in left intraparietal sulcus, one anterior and one posterior (Table 3, Fig. 3A).

The double “mismatch” versus correct condition contrast showed an area in the left inferior frontal cortex and an area in the precentral cortex to be differentially activated (Table 3; Fig. 4). The fact that precentral cortex is also activated in this comparison, may seem to be in contrast to the findings from the ROI analysis in BA 6, reported above. In the ROI analysis, only the gesture “mismatch” condition led to significantly increased activation compared with the correct condition. The small region found activated to the double “mismatch” condition in the whole-brain analysis is part of the ROI used to test differential activation of BA 6. The fact that no difference is found in the ROI analysis of BA 6 between double “mismatch” and correct condition is probably because the extent of the reaction of premotor cortex is more limited to the double “mismatch” condition than to the gesture “mismatch” condition. In other words, modulation of premotor cortex seems to be much more robust in the gesture “mismatch” condition than in the double “mismatch” condition.

In none of the contrasts significant differences in the opposite direction (i.e., correct > “mismatch”) were found.

Next, a conjunction analysis, performed by a test for independently significant effects as in a logical AND (Nichols et al. 2005), was performed to test for regions involved in both the language “mismatch” as well as the gesture “mismatch” conditions. In this way, common (overlapping) activations for both processes could be defined. This analysis tested for areas activated significantly stronger in both the language “mismatch” versus correct condition contrast and the gesture “mismatch” versus correct condition contrast (G+L- vs. G+L+ \cap G-L+ vs. G+L+). One area in the left inferior frontal cortex was significantly activated in this conjunction (overlap in Fig. 3A).

Table 3

Activations in the whole-brain analysis

Contrast	Centre coordinates			Region	<i>T</i> (max)	Number of voxels (1 × 1 × 1 mm)
	<i>x</i>	<i>y</i>	<i>z</i>			
Language mismatch versus correct	-43	12	24	Left inferior frontal sulcus	5.19	7326
	-33	-65	35	Left intraparietal sulcus (posterior)	4.33	1591
	-52	-50	4	Left superior temporal sulcus	4.83	2374
Gesture mismatch versus correct	-46	29	23	Left inferior frontal gyrus/sulcus	4.48	1651
	-32	-46	31	Left intraparietal sulcus (anterior)	5.93	2531
	-19	-63	32	Left intraparietal sulcus (posterior)	5.37	2290
Double mismatch versus correct	-55	17	24	Left inferior frontal sulcus	5.96	1586
	-52	-6	49	Left precentral sulcus	4.36	1145
Conjunction of G+L- versus G-L- and G-L+ versus G-L-	38	-54	-13	Right fusiform gyrus	4.82	1946

Note: Regions that were significantly activated in the whole-brain random effects group analysis ($t(15) > 3.5$, $P < 0.05$, corrected). Displayed are the contrasts, the center coordinates in stereotaxic space (Talairach and Tournoux 1988), a description of the region, the *T* value of the maximally activated voxel, and the number of significant voxels.

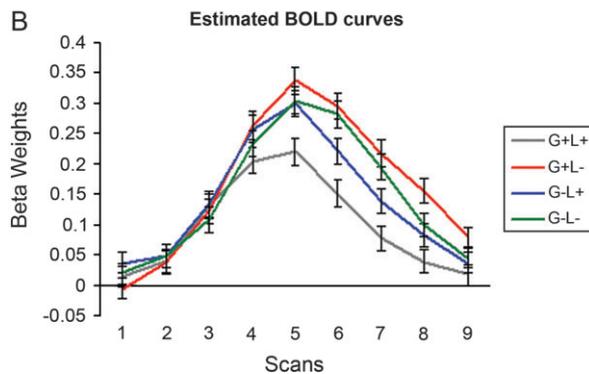
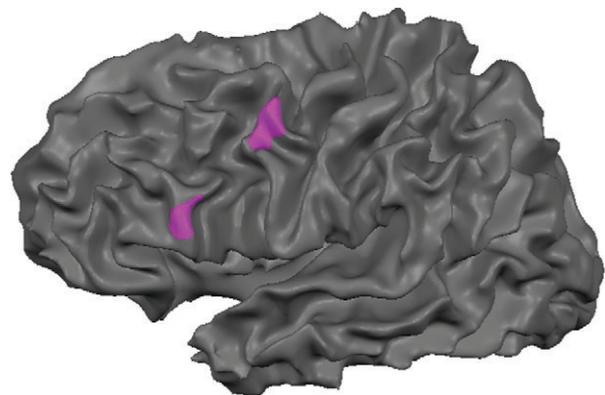
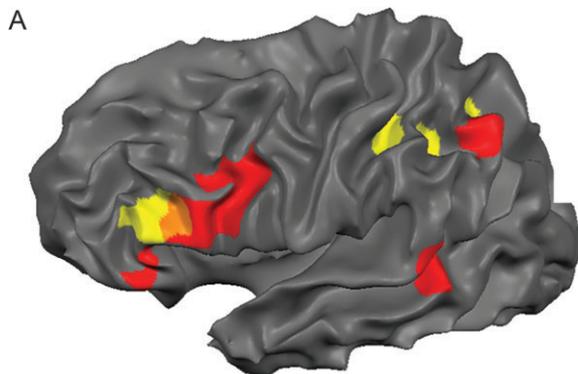


Figure 3. Gesture and speech in a sentence context. (A) Significant activations in the whole-brain random effects analysis for the language “mismatch” versus correct (red) and the gesture “mismatch” versus correct (yellow) conditions. Note the overlap in inferior frontal cortex (BA 45, [*x y z*] [-46 23 25] (Talairach and Tournoux 1988)). Maps are thresholded at $t(15) > 3.5$, $P < 0.05$ (corrected) and projected onto the cortical sheet of one of the participants. No activations were found in the right hemisphere. (B) BOLD curves from the activated regions in left inferior frontal cortex (center coordinates [*x y z*] [-43 11 26] (Talairach and Tournoux 1988)). The curves were created by estimating each time point after modeling an event by 9 subsequent stick functions (Dale 1999; Miezin et al. 2000). This region is also activated in the correct condition (gray line) but more so in reaction to a semantic “mismatch” (red, blue and green lines). The time scale on the *x*-axis is in TRs (“scans”), one scan is 2230 ms.

Comparison with a cytoarchitectonic probability map of BA 45 (Amunts et al. 1999; Eickhoff et al. 2005), showed that 82% of the voxels in this region were part of BA 45 with its center of gravity ([*x y z*] [-46 23 25] (Talairach and Tournoux 1988)) having a probability of 60% to be part of BA 45. When the statistical threshold was lowered for informal visual inspection, the area of overlap became much bigger. This confirms that the

Figure 4. Significant activations in the whole-brain random effects analysis in the double “mismatch” versus correct condition contrast. The inferior frontal area and the precentral area in purple were more strongly activated by the double “mismatch” (G-L-) condition than by the correct condition (G+L+). Map is thresholded at $t(15) > 3.5$, $P < 0.05$ (corrected) and projected onto the cortical sheet of one of the participants. No activations were found in the right hemisphere.

activations in left inferior frontal cortex displayed in Figure 3(A) are not 2 distinct areas slightly overlapping but are part of the same region being activated in both conditions.

To test for activations specifically obtained for the local mismatch in the gesture and language “mismatch” conditions, a conjunction analysis (Nichols et al. 2005) was performed, testing for regions activated in both the G+L- versus G-L- and the G-L+ versus G-L- contrasts. This analysis was preferred over an analysis contrasting the 2 conditions with a local mismatch versus the 2 conditions with a local match because the latter analysis would contain a confound between overall sentence anomaly (absent for the correct condition) and local match (present for the correct and double “mismatch” conditions). To avoid such a confound, we contrasted the local mismatch conditions (language “mismatch,” G+L-; gesture “mismatch,” G-L+) against the double “mismatch” condition (G+L- vs. G-L- ∩ G-L+ vs. G-L-). In this way all conditions were semantically anomalous, so that the only remaining difference was a local mismatch (G+L-; G-L+) versus a local match (G-L-). One area in the right fusiform gyrus (Table 3; Supplementary Fig. S4) was found to be significantly more activated in the conditions with a local mismatch (G+L- and G-L+) compared with the condition with a local match (G-L-). (However, at the more liberal voxel-threshold, intraparietal areas were activated bilaterally, as is already suggested by Fig. 2C).

At the end of the scanning session, participants were extensively debriefed. All participants were able to describe the manipulations in the materials and could provide examples of specific trials. All participants were aware of the cases in which language and/or gestures did not fit well into the preceding sentence context. Moreover, they were aware of both language and gesture “mismatches” to an approximately equal extent, indicating that both language and gesture information had been in the focus of attention.

Discussion

The main goal of this study was to investigate the neural integration of semantic information conveyed through language (speech) and action (gestures) within a sentence context. The results show that both action and language recruit overlapping parts of left inferior frontal cortex, specifically BA 45. That is, this region is modulated by an increase in the semantic load of simultaneously presented information from the speech and action domains. Additionally, premotor cortex is modulated by the semantic processing of actions within a language context. These results are in line with accounts hypothesizing a link between language and action systems. Furthermore, the fact that we found overlapping areas provides neural evidence for claims that speech and gesture are closely linked in language comprehension (McNeill 1992, 2000; Goldin Meadow 2003; Kendon 2004).

The involvement of left inferior frontal cortex in integrating semantic information from both the action and language domains is consistent with a theory of language comprehension in which the left inferior frontal cortex serves as the general (i.e., not domain-specific) unification site for language comprehension (Hagoort 2003; Hagoort 2005b). During unification, current information is integrated into an unfolding representation of multiword utterances. In the case of a semantic “mismatch,” integration of the anomalous information is harder, resulting in an activation increase. The blood oxygen level-dependent (BOLD) curves of the left inferior frontal cortex (Fig. 3B) show that this increased activation does not reflect a reaction to a semantic mismatch as such but that the area is also activated in the processing of correct sentences. This lends credibility to the idea that Broca’s area and adjacent cortex, more in particular BA 45 and 47, is the semantic unification site for language comprehension. Its activation in studies using the mismatch paradigm is no artifact of the materials used but truly reflects increased semantic load. Moreover, most of our items in the “mismatching” conditions were not straightforward semantic violations but contained only semantically less expected verbs and/or gestures given the semantics of the preceding sentence context. This is in line with a recent study (Rodd et al. 2005) in which stimulus materials contained no semantic violations whatsoever. Yet, sentences containing semantically ambiguous words activated left inferior frontal cortex stronger than sentences containing nonambiguous words. Our results are also in accordance with the established finding in the ERP literature that both straightforward semantic anomalies as well as subtle manipulations in semantic integration processes lead to the same ERP effect, namely the N400 effect (Kutas and Hillyard 1980; Kutas and Hillyard 1984; Hagoort and Brown 1994; Kutas and Van Petten 1994). Moreover, in an ERP study conducted in our lab (Özyürek et al. in press) with the same materials as used in this study it was found that ERPs time

locked to the critical verbs and gestures elicited an N400 effect in all 3 “mismatch” conditions. Furthermore, the latency and amplitude of these N400 effects were similar. This is independent evidence for the claim that the effects reported here reflect semantic unification as indexed by the N400 effects. It furthermore shows that processing of gestures versus critical verbs in relation to previous context does not involve different processing strategies, which is consistent with the debriefing reports of the participants in our study.

Based on the results for the language and gesture “mismatch” conditions, one could argue that the activation of the left inferior frontal cortex is not due to the increased semantic integration load with respect to the preceding sentence context but instead to the local mismatch between the simultaneously occurring critical verb and the gesture (e.g., the verb hit and the gesture write). However, this interpretation is challenged by the fact that we observed the same area activated in the double “mismatch” condition (Fig. 4, Table 3), in which verb and gesture match with respect to each other and the “mismatch” is solely in relation to the preceding context.

These findings are in one important aspect different from what has been reported for the audiovisual integration of speech and lip movements, where mismatching information has been found to result in a reduced activation compared with matching information (Calvert 2001). This difference might be due to the fact that for the integration of speech and lip movements, in the matching condition the auditory and visual input converge on a common form representation in memory (e.g., a particular syllable) that as a result gets more strongly activated than in a mismatching condition. In our study, the gesture and speech signal had to be integrated in a sentence-level semantic representation that is not available in memory but has to be constructed online. This semantic integration process is more strongly taxed when the integration load increases leading to higher activation levels in the “mismatch” conditions.

Overall, our findings are compatible with other studies on language-related integration processes. It was recently shown that an area in left inferior frontal cortex partly overlapping with the region reported here is involved in the integration of both semantic and world knowledge information during reading (Hagoort et al. 2004). The present study replicates the role of this area in integrating information into a prior sentence context but this time with spoken input. Within the domain of language, this area seems to operate independent of input modality (reading vs. speech). Importantly, the data presented here convincingly demonstrate that unification in the left inferior frontal cortex during language comprehension is not domain specific. The integration of semantic information conveyed through the action domain also recruits this area. A dorsal to ventral parcellation of inferior frontal cortex into distinct subregions performing different core functions within the language domain, such as phonological, semantic, and syntactic processing, has been proposed (Poldrack et al. 1999; Bookheimer 2002; Vigneau et al. 2006). Note that the location of the activation to both gesture and language conditions in this study is in line with the location of the semantic component within left inferior frontal cortex.

Finally, our results also show overlap with sign language comprehension with regard to the involvement of left inferior frontal and temporal cortices (Neville et al. 1998; MacSweeney et al. 2002, 2004; MacSweeney et al. 2006), despite the differences in linguistic properties of signs compared with gestures.

This study also sheds light on the semantic modulation of the action recognition system. The fact that we found a context-dependent modulation of premotor cortex (BA 6) has important implications for the role of this area in action observation. A large number of studies show activation of premotor areas by action observation (Hari et al. 1998; Nishitani and Hari 2000; Buccino et al. 2001; Jeannerod 2001; Rizzolatti et al. 2001; Grezes et al. 2003; Costantini et al. 2005). This has been interpreted as evidence for the existence of an action recognition system in humans, comparable with the mirror neuron system in monkeys, in which similar neural activations exist during action observation and action execution (Jeannerod 2001; Rizzolatti et al. 2001; Nishitani et al. 2005). Premotor activation is also found when stimuli are meaningless actions (Fadiga et al. 1995; Decety et al. 1997), point light displays (Saygin et al. 2004), biologically impossible actions (Costantini et al. 2005), and in motor imagery (Schubotz and von Cramon 2004; de Lange et al. 2005). Together, these findings suggest that the activation of the premotor cortex (BA 6) is automatic and occurs to the observation of any type of action. Here we show that although possibly automatic, activation of premotor cortex is influenced by semantic information from speech. That is, premotor cortex is directly sensitive to the semantic context in which an action occurs, possibly through top-down modulations of motor representations by higher order cortical areas. Future work is needed to investigate the specific neural dynamics of this interaction.

One interesting finding with regard to the modulation of premotor cortex is that this area was modulated in a more robust way in the gesture “mismatch” than in the double “mismatch” condition, even though gesture information was harder to integrate to previous context in both conditions. In the ROI analysis, the premotor cortex was found to be activated only in the gesture “mismatch” condition. However, in the whole-brain analysis, a small region of precentral gyrus was found activated to the double “mismatch” condition as well. Apparently, in the double “mismatch” condition premotor areas are activated to some extent but much less robustly than for the gesture “mismatch” condition. One could speculate that the more robust activation in the gesture “mismatch” is due to the fact that in this condition there is an additional local mismatch between the co-occurring verb and gesture, which is not present in the double “mismatch” condition.

Given their commonly observed role in action observation, it is tempting to interpret the activations of intraparietal regions in the gesture “mismatch” condition in the whole-brain analysis in a similar vein as the findings in BA 6. However, the activation of intraparietal areas was not specific to the gesture condition but was also found in reaction to the language “mismatch” condition, albeit in a slightly different location. Therefore, we interpret these activations in another, more parsimonious way. That is, both these conditions might lead to increased spatial attention, a process in which (intra)parietal regions are known to also be involved (Corbetta and Shulman 2002). The non-specific nature of these activations (not in response to one particular condition) strengthens this explanation.

Finally, a left superior temporal activation was seen in the language “mismatch” condition. This finding is compatible with previous studies of semantic aspects of sentence processing (Ni et al. 2000; Kuperberg et al. 2003). Presumably this activation reflects the interaction between context and the retrieval of lexical-semantic information.

In summary, our results reveal 2 important aspects of the relations between language and action systems. One is that high-level neural integration of semantic information into a context is not domain specific and takes place in Broca’s area. When understanding a sentence, the brain does not restrict itself to language information alone but also integrates semantic action information conveyed through cospeech gestures into the preceding message context. Both action and language semantics constrain the interpretation domain simultaneously, and by recruiting the neural contribution of left inferior frontal cortex. This opens the interesting possibility that neural processing in language comprehension involves the incorporation of information in a single unification space coming from a broader range of cognitive domains than thought so far.

Different proposals on the role of inferior frontal cortex have been put forward in recent years, such as selection among competing alternatives (Thompson-Schill et al. 1997), controlled semantic retrieval (Wagner et al. 2001), or both (Badre et al. 2005). Such views are not inconsistent with our account because selection is a necessary aspect of unification, as we have argued elsewhere and as is specified in explicit computational models of unification (Vosse and Kempen 2000; Hagoort 2005a). How the role of left inferior frontal cortex is best characterized if one wants to cover all findings available in the literature, is an open question. One possibility is that seemingly conflicting findings regarding inferior frontal cortex functioning can be subsumed under the heading of one underlying common process, such as the “regulation of mental activity” (Thompson-Schill et al. 2005). However, from the perspective of the cortex as a dynamically changing system of large-scale distributed functional networks (Mesulam 1990; Mesulam 1998; Fuster 2003), it is conceivable that higher order cortices do not perform one function but play different roles in different networks depending upon the nature of input and task. Along these lines, we do not claim that the sole function of this cortical area is to be involved in semantic unification. Whether different functional accounts of inferior frontal cortex can be grouped under one heading (Thompson-Schill et al. 2005) or that qualitatively different functions can coexist within one part of cortex is an important question for future research. In any case, accounts of inferior frontal cortex as playing a role in selection (Thompson-Schill et al. 1997; Badre et al. 2005) are compatible with our interpretation that this region contributes to unification processes.

The second important finding of our study is that contextual information from the language domain can influence parts of the motor system. This adds to the growing insight that cognitive modulation of areas at a lower level in the cortical hierarchy appears to be an important principle in the neural architecture of human cognition (see also de Araujo et al. 2005).

In conclusion, we have shown that a classical language area, Broca’s area, can be modulated by action processing as well as that a classical action area, premotor cortex, can be modulated by the language context in which actions are embedded. These findings provide support for the claims that in real life speech and action are often tightly interconnected (McNeill 1992; McNeill 2000; Goldin Meadow 2003; Kita and Özyürek 2003; Kendon 2004), and that there are close links between the action and language systems (Rizzolatti and Arbib 1998; Nishitani et al. 2005). Many aspects of their neural interplay remain to be unraveled but this study provides a first insight into the neural integration of language and action information.

Supplementary Data

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>.

Notes

This research was supported by a grant from the Netherlands Organization for Scientific Research, 051.02.040. We thank Ivan Toni, Michael Coles, Sotaro Kita, Floris de Lange, and Marieke Schölvincq for their comments on an earlier version of our paper and Paul Gaalman for assistance during the scanning sessions. Petra van Alphen is acknowledged for lending her voice to the sentence materials. We thank the 3 anonymous reviewers for their insightful comments.

Conflict of Interest: None declared.

Address correspondence to F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands. Email: roel.willems@fcdonders.ru.nl.

References

- Aboitiz F, Garcia R. 1997. The evolutionary origin of the language areas in the human brain: a neuroanatomical perspective. *Brain Res Rev.* 25(3):381-396.
- Aboitiz F, Garcia RR, Bosman C, Brunetti E. 2006. Cortical memory mechanisms and language origins. *Brain Lang.* 98(1):40-56.
- Amunts K, Schleicher A, Burgel U, Mohlberg H, Uylings HB, Zilles K. 1999. Broca's region revisited: cytoarchitecture and intersubject variability. *J Comp Neurol.* 412(2):319-341.
- Arbib MA. 2005. From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behav Brain Sci.* 28(2):105-124.
- Badre D, Poldrack RA, Pare-Blagoev EJ, Insler RZ, Wagner AD. 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron.* 47(6):907-918.
- Beattie G, Shovelton H. 1999. Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *J Lang Soc Psychol.* 18(4):438-462.
- Bernardis P, Gentilucci M. 2006. Speech and gesture share the same communication system. *Neuropsychologia.* 44(2):178-190.
- Bookheimer S. 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci.* 25:151-188.
- Bosman C, Lopez V, Aboitiz F. 2005. Sharpening Occam's razor: is there need for a hand-signing stage prior to vocal communication? *Behav Brain Sci.* 28(2):128-129.
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Seitz RJ, Zilles K, Rizzolatti G, Freund HJ. 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur J Neurosci.* 13(2):400-404.
- Buccino G, Vogt S, Ritzl A, Fink GR, Zilles K, Freund HJ, Rizzolatti G. 2004. Neural circuits underlying imitation learning of hand actions: an event-related fMRI study. *Neuron.* 42(2):323-334.
- Butterworth G, Shovelton H. 1978. Gesture and silence as indicators of planning in speech. In: Campbell N, Smith P, editors. *Recent advances in the psychology of language*. New York: Plenum. p. 347-360.
- Calvert GA. 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex.* 11(12):1110-1123.
- Cassell J, McNeill D, McCullough KE. 1999. Speech-gesture mismatches: evidence for one underlying representation of linguistic and non-linguistic information. *Pragmatics Cogn.* 7(1):1-34.
- Church RB, Goldin Meadow S. 1986. The mismatch between gesture and speech as an index of transitional knowledge. *Cognition.* 23(1):43-71.
- Clark HH. 1996. *Using language*. New York, NY: Cambridge University Press.
- Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci.* 3(3):201-215.
- Corina DP, Knapp H. 2006. Sign language processing and the mirror neuron system. *Cortex.* 42(4):529-539.
- Costantini M, Galati G, Ferretti A, Caulo M, Tartaro A, Romani GL, Aglioti SM. 2005. Neural systems underlying observation of humanly impossible movements: an fMRI study. *Cereb Cortex.* 15(11):1761-1767.
- Dale AM. 1999. Optimal experimental design for event-related fMRI. *Hum Brain Mapp.* 8(2-3):109-114.
- de Araujo IE, Rolls ET, Velazco MI, Margot C, Cayeux I. 2005. Cognitive modulation of olfactory processing. *Neuron.* 46(4):671-679.
- de Lange FP, Hagoort P, Toni I. 2005. Neural topography and content of movement representations. *J Cogn Neurosci.* 17(1):97-112.
- Decety J, Grezes J, Costes N, Perani D, Jeannerod M, Procyk E, Grassi F, Fazio F. 1997. Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain.* 120(Pt 10):1763-1777.
- Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage.* 25(4):1325-1335.
- Emmorey K. 2006. The role of Broca's area in sign language. In: Grodzinsky Y, Amunts K, editors. *Broca's region*. New York: Oxford University Press. p. 169-184.
- Fadiga L, Fogassi L, Pavesi G, Rizzolatti G. 1995. Motor facilitation during action observation: a magnetic stimulation study. *J Neurophysiol.* 73(6):2608-2611.
- 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(5):636-647.
- Friederici AD, Rüschemeyer S, Hahne A, Fiebach CJ. 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb Cortex.* 13:170-177.
- Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R. 1998. Event-related fMRI: characterizing differential responses. *Neuroimage.* 7(1):30-40.
- Fuster JM. 2003. *Cortex and mind*. New York: Oxford University Press.
- Gallagher HL, Frith CD. 2004. Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. *Neuropsychologia.* 42(13):1725-1736.
- Goldin Meadow S. 2003. *Hearing gesture: how our hands help us think*. Cambridge, MA: Belknap Press of Harvard University Press.
- Goldin Meadow S, Kim S, Singer M. 1999. What the teacher's hands tell the student's mind about math. *J Educ Psychol.* 91(4):720-730.
- Goldin Meadow S, Momeni Sandhofer C. 1999. Gestures convey substantive information about a child's thoughts to ordinary listeners. *Dev Sci.* 2(1):67-74.
- Grezes J, Armony JL, Rowe J, Passingham RE. 2003. Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. *Neuroimage.* 18(4):928-937.
- Grezes J, Costes N, Decety J. 1999. The effects of learning and intention on the neural network involved in the perception of meaningless actions. *Brain.* 122(Pt 10):1875-1887.
- Hagoort P. 2003. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *Neuroimage.* 20(Suppl 1):S18-S29.
- Hagoort P. 2005a. Broca's complex as the unification space for language. In: Cutler A, editor. *Twenty first century psycholinguistics: four cornerstones*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers. p. 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, editors. *Perspectives in sentence processing*. Hillsdale, NJ, UK: Lawrence Erlbaum Associates. p. 45-80.
- Hagoort P, Hald L, Bastiaansen M, Petersson KM. 2004. Integration of word meaning and world knowledge in language comprehension. *Science.* 304(5669):438-441.
- Hamzei F, Rijntjes M, Dettmers C, Glauche V, Weiller C, Büchel C. 2003. The human action recognition system and its relationship to Broca's area: an fMRI study. *Neuroimage.* 19(3):637-644.

- Hari R, Forss N, Avikainen S, Kirveskari E, Salenius S, Rizzolatti G. 1998. Activation of human primary motor cortex during action observation: a neuromagnetic study. *Proc Natl Acad Sci USA*. 95(25):15061-15065.
- Holle H, Gunter TC. 2006. Dynamic co-speech gestures may be used as disambiguation cues: ERP evidence. *J Cogn Neurosci (Suppl)*. 18:221.
- Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G. 1999. Cortical mechanisms of human imitation. *Science*. 286(5449):2526-2528.
- Jeannerod M. 2001. Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage*. 14(1 Pt 2):S103-S109.
- Josephs O, Turner R, Friston K. 1997. Event-related fMRI. *Hum Brain Mapp*. 5:243-248.
- Kelly SD, Barr DJ, Church RB, Lynch K. 1999. Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. *J Mem Lang*. 40(4):577-592.
- Kelly SD, Church RB. 1998. A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Dev*. 69(1):85-93.
- Kelly SD, Kravitz C, Hopkins M. 2004. Neural correlates of bimodal speech and gesture comprehension. *Brain Lang*. 89(1):253-260.
- Kendon A. 2004. *Gesture: visible action as utterance*. Cambridge, UK: Cambridge University Press.
- 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.
- Koski L, Wohlschläger A, Bekkering H, Woods RP, Dubeau MC, Mazziotta JC, Iacoboni M. 2002. Modulation of motor and premotor activity during imitation of target-directed actions. *Cereb Cortex*. 12(8):847-855.
- Krauss RM, Morrel Samuels P, Colasante C. 1991. Do conversational hand gestures communicate? *J Pers Soc Psychol*. 61(5):743-754.
- Kuperberg GR, Holcomb PJ, Sitnikova T, Greve D, Dale AM, Caplan D. 2003. Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *J Cogn Neurosci*. 15(2):272-293.
- Kuperberg GR, McGuire PK, Bullmore ET, Brammer MJ, Rabe-Hesketh S, Wright IC, Lythgoe DJ, Williams SC, David AS. 2000. Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: an fMRI study. *J Cogn Neurosci*. 12(2):321-341.
- Kutas M, Hillyard SA. 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*. 207(4427):203-205.
- Kutas M, Hillyard SA. 1984. Brain potentials during reading reflect word expectancy and semantic association. *Nature*. 307(5947):161-163.
- Kutas M, Van Petten CK. 1994. Psycholinguistics electrified: event-related brain potential investigations. In: Gernsbacher MA, editor. *Handbook of psycholinguistics*. San Diego, CA: Academic Press. p. 83-143.
- MacSweeney M, Campbell R, Woll B, Brammer MJ, Giampietro V, David AS, Calvert GA, McGuire PK. 2006. Lexical and sentential processing in British Sign Language. *Hum Brain Mapp*. 27(1):63-76.
- MacSweeney M, Campbell R, Woll B, Giampietro V, David AS, McGuire PK, Calvert GA, Brammer MJ. 2004. Dissociating linguistic and nonlinguistic gestural communication in the brain. *Neuroimage*. 22(4):1605-1618.
- MacSweeney M, Woll B, Campbell R, McGuire PK, David AS, Williams SC, Suckling J, Calvert GA, Brammer MJ. 2002. Neural systems underlying British Sign Language and audio-visual English processing in native users. *Brain*. 125(Pt 7):1583-1593.
- McNeill D. 1992. *Hand and mind: what gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- McNeill D. 2000. *Language and gesture*. Cambridge, UK: Cambridge University Press.
- McNeill D, Cassell J, McCullough KE. 1994. Communicative effects of speech-mismatched gestures. *Res Lang Soc Interact*. 27(3):223-237.
- Melinger A, Levelt WJM. 2004. Gesture and the communicative intention of the speaker. *Gesture*. 4(2):119-141.
- Mesulam MM. 1990. Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Ann Neurol*. 28(5):597-613.
- Mesulam MM. 1998. From sensation to cognition. *Brain*. 121(Pt 6):1013-1052.
- Miezin FM, Maccotta L, Ollinger JM, Petersen SE, Buckner RL. 2000. Characterizing the hemodynamic response: effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing. *Neuroimage*. 11(6 Pt 1):735-759.
- Molnar-Szakacs I, Iacoboni M, Koski L, Mazziotta JC. 2005. Functional segregation within pars opercularis of the inferior frontal gyrus: evidence from fMRI studies of imitation and action observation. *Cereb Cortex*. 15(7):986-994.
- Morrel Samuels P, Krauss RM. 1992. Word familiarity predicts temporal asynchrony of hand gestures and speech. *J Exp Psychol Learn Mem Cogn*. 18(3):615-622.
- Nakamura A, Maess B, Knosche TR, Gunter TC, Bach P, Friederici AD. 2004. Cooperation of different neuronal systems during hand sign recognition. *Neuroimage*. 23(1):25-34.
- Neville HJ, Bavelier D, Corina D, Rauschecker J, Karni A, Lalwani A, Braun A, Clark V, Jezzard P, Turner R. 1998. Cerebral organization for language in deaf and hearing subjects: biological constraints and effects of experience. *Proc Natl Acad Sci USA*. 95(3):922-929.
- Newman AJ, Bavelier D, Corina D, Jezzard P, Neville HJ. 2002. A critical period for right hemisphere recruitment in American Sign Language processing. *Nat Neurosci*. 5(1):76-80.
- Ni W, Constable RT, Mencl WE, Pugh KR, Fullbright RK, Shaywitz SE, Shaywitz BA, Gore JC, Shankweiler D. 2000. An event-related neuroimaging study distinguishing form and content in sentence processing. *J Cogn Neurosci*. 12(1):120-133.
- Nichols T, Brett M, Andersson J, Wager T, Poline JB. 2005. Valid conjunction inference with the minimum statistic. *Neuroimage*. 25(3):653-660.
- Nishitani N, Hari R. 2000. Temporal dynamics of cortical representation for action. *Proc Natl Acad Sci USA*. 97(2):913-918.
- Nishitani N, Schürmann M, Amunts K, Hari R. 2005. Broca's region: from action to language. *Physiology*. 20:60-69.
- Oldfield RC. 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 RM, Kita S, Hagoort P. Forthcoming. On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J Cogn Neurosci*.
- Petersson KM, Forkstam C, Ingvar M. 2004. Artificial syntactic violations activate Broca's region. *Cogn Sci*. 28:383-407.
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JD. 1999. Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*. 10(1):15-35.
- Pulvermüller F. 2005. Brain mechanisms linking language and action. *Nat Rev Neurosci*. 6(7):576-582.
- Rizzolatti G, Arbib MA. 1998. Language within our grasp. *Trends Neurosci*. 21(5):188-194.
- Rizzolatti G, Fogassi L, Gallese V. 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci*. 2(9):661-670.
- Rodd JM, Davis MH, Johnsrude IS. 2005. The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb Cortex*. 15(8):1261-1269.
- Ruschemeyer SA, Zysset S, Friederici AD. 2006. Native and non-native reading of sentences: an fMRI experiment. *Neuroimage*. 31(1):354-365.
- Saygin AP, Wilson SM, Hagler DJ, Jr, Bates E, Sereno MI. 2004. Point-light biological motion perception activates human premotor cortex. *J Neurosci*. 24(27):6181-6188.
- Schubotz RI, von Cramon DY. 2004. Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. *J Neurosci*. 24(24):5467-5474.

- Singer MA, Goldin Meadow S. 2005. Children learn when their teacher's gestures and speech differ. *Psychol Sci.* 16(2):85-89.
- Talairach J, Tournoux P. 1988. Co-planar stereotaxic atlas of the human brain. New York: Thieme Medical.
- Thompson LA, Massaro DW. 1986. Evaluation and integration of speech and pointing gestures during referential understanding. *J Exp Child Psychol.* 42(1):144-168.
- Thompson LA, Massaro DW. 1994. Children's integration of speech and pointing gestures in comprehension. *J Exp Child Psychol.* 57(3):327-354.
- Thompson-Schill SL, Bedny M, Goldberg RF. 2005. The frontal lobes and the regulation of mental activity. *Curr Opin Neurobiol.* 15(2): 219-224.
- Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. 1997. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc Natl Acad Sci USA.* 94(26):14792-14797.
- Vigneau M, Beaucousin V, Herve PY, 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.
- Vosse T, Kempen G. 2000. Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and lexicalist grammar. *Cognition.* 75(2):105-143.
- Wagner AD, Pare-Blagoev EJ, Clark J, Poldrack RA. 2001. Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron.* 31(2):329-338.
- Wu YC, Coulson S. 2005. Meaningful gestures: electrophysiological indices of iconic gesture comprehension. *Psychophysiology.* 42(6):654-667.
- Xiong J, Rao S, Jerabek P, Zamarripa F, Woldorff M, Lancaster J, Fox PT. 2000. Intersubject variability in cortical activations during a complex language task. *Neuroimage.* 12(3):326-339.