

# The Differentiation of Iconic and Metaphoric Gestures: Common and Unique Integration Processes

Benjamin Straube,<sup>1,2\*</sup> Antonia Green,<sup>2,3</sup> Bianca Bromberger,<sup>1</sup>  
and Tilo Kircher<sup>2</sup>

<sup>1</sup>Department of Neurology and the Center for Cognitive Neuroscience, The University of Pennsylvania, Philadelphia, Pennsylvania

<sup>2</sup>Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Marburg, Germany

<sup>3</sup>Department of Psychiatry and Psychotherapy, Section of Experimental Psychopathology, RWTH Aachen University, Aachen, Germany

---

**Abstract:** Recent research on the neural integration of speech and gesture has examined either gesture in the context of concrete [iconic (IC) gestures] or abstract sentence content [metaphoric (MP) gestures]. However, there has not yet been a direct comparison of the processing of both gesture types. This study tested the theory that left posterior temporal and inferior frontal brain regions are each uniquely involved in the integration of IC and MP gestures. During fMRI-data acquisition, participants were shown videos of an actor performing IC and MP gestures and associated sentences. An isolated gesture (G) and isolated sentence condition (S) were included to separate unimodal from bimodal effects at the neural level. During IC conditions, we found increased activity in the left posterior middle temporal gyrus and its right hemispheric homologue. The same regions in addition to the left inferior frontal gyrus (IFG) were activated during MP conditions in contrast to the isolated conditions (G&S). These findings support the hypothesis that there are distinct integration processes for IC and MP gestures. In line with recent claims of the semantic unification theory, there seems to be a division between perceptual-matching processes within the posterior temporal lobe and higher-order relational processes within the IFG. *Hum Brain Mapp* 32:520–533, 2011. © 2010 Wiley-Liss, Inc.

**Key words:** coverbal gestures; metaphoric gestures; iconic gestures; integration; unification

---

## INTRODUCTION

The field of neuroscience is becoming increasingly interested in interpersonal communication and, in particular, the interaction between speech and gesture. However, there are discrepant findings and differing theories about the neural substrates that underlie this relationship. This study directly compared the processing of two important coverbal gestures types to disentangle the roles of specific brain regions in the integration of speech and gesture.

Different types of gestures vary in their relation to language [McNeill, 1992, 2005]. Iconic (IC) and metaphoric (MP) gestures both illustrate spoken sentences, but do so in unique ways. IC gestures refer to the concrete content

---

Contract grant sponsor: Interdisciplinary Center for Clinical Research "BIOMAT", Aachen, Germany; Contract grant number: IZKF VV N68.

\*Correspondence to: Benjamin Straube, Department of Psychiatry und Psychotherapy, Philipps-University Marburg; Rudolf-Bultmann Str. 8, 35039 Marburg, Germany.  
E-mail: [straubeb@med.uni-marburg.de](mailto:straubeb@med.uni-marburg.de)

Received for publication 16 September 2009; Revised 12 January 2010; Accepted 8 February 2010

DOI: 10.1002/hbm.21041

Published online 13 May 2010 in Wiley Online Library ([wileyonlinelibrary.com](http://wileyonlinelibrary.com)).

of sentences, whereas MP gestures illustrate abstract information. For example, in the sentences “He gets down to business” (drop of the hand) or “The politician builds a bridge to the next topic” (depicting an arch with the hand), abstract content is illustrated by a MP gesture. However, the same gestures can be IC (drop of the right hand or depicting an arch with the right hand) when paired with the sentences “The man goes down the hill” or “There is a bridge over the river” as they illustrate concrete physical features of the world.

There is increasing evidence that language and gesture comprehension rely on partially overlapping brain networks (for a review, see Willems and Hagoort [2007]). Recent fMRI investigations have focused on the interaction between speech and gesture, using different types of speech-associated gestures such as beat gestures, IC gestures and MP gestures [Green et al., 2009; Holle et al., 2008; Hubbard et al., 2009; Kircher et al., 2009; Straube et al., 2009; Willems et al., 2007]. Hubbard et al. [2009] showed that beat gestures (hand gestures that mark speech prosody) modulate activity in the auditory cortex during speech processing. These beat gestures convey rhythmic structural information rather than semantic information. In contrast, IC and MP gestures contain sentence-related semantic information. Holle et al. [2008] investigated IC coverbal gestures through a disambiguation paradigm. They showed that the posterior superior temporal sulcus (STS) is activated when gestures clarify ambiguous sentences, suggesting that the left STS plays a specific role in the integration of speech and gesture. However, Willems et al. [2007] showed in a mismatch paradigm that the left inferior frontal gyrus (IFG) is activated for both speech and gesture mismatches. They concluded that the increased activation of left inferior frontal regions reflects the increased integration load and reasoned that the IFG is responsible for speech and gesture integration. Finally, Green et al. [2009] replicated the finding that the left IFG and its right hemispheric homologue are activated for incongruent versus congruent IC coverbal gesture processing. However, in comparison with the processing of gestures in the context of a foreign language, they demonstrated that the left posterior temporal lobes appear to be involved in the semantic integration processes of speech and gesture. Taken together, these findings suggest that the left posterior temporal lobe is involved in the integration of speech and IC gestures, whereas the left or bilateral IFG is involved when speech and gestures are unrelated or mismatched.

In contrast to the attention paid to IC gesture processing, there have been relatively few studies that have investigated the integration of MP gestures. However, the left posterior temporal lobe, the premotor cortex, and the IFG are found to be more active during conditions of combined speech and gestures in contrast to isolated conditions [Kircher et al., 2009]. Activity in these areas is also correlated with performance on subsequent memory tasks that are also indicative of integration processes [Straube

et al., 2009]. Assuming that the “integration load” is high for gestures in the context of speech with abstract content (as the gestures must be interpreted), these data agree fairly well with the theory that the left posterior temporal lobe is involved in general integration processes, whereas the left IFG is involved whenever the bimodal processing load is high.

Until now, few studies have investigated the processing of different coverbal gesture types. In a recent study, Willems et al. [2009] investigated IC coverbal gestures and pantomimes. In this study, speech and IC coverbal gestures are always presented together, and IC gestures cannot be unambiguously understood without speech. However, pantomimes are not necessarily produced together with speech and can be easily understood without speech. In this study, speech is presented with these two types of communicative hand actions in matching or mismatching combinations to manipulate the semantic integration load. The left and right pSTS/MTG was only involved in the semantic integration of speech and pantomimes. However, the left IFG was involved in the integration of speech and coverbal gestures as well as of speech and pantomimes. It was suggested 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 the LIFG is better characterized as the online construction of a new and unified representation of the input streams. Nonetheless, the study does not provide much information about the integration of gestures in the context of abstract versus concrete sentences. The study generally suggests that pSTS/MTG and LIFG are differentially involved in multimodal integration, depending on the semantic relationship between the input streams.

Within the framework of the semantic unification theory [Hagoort et al., 2009], there is a possible division of labor between inferior frontal and superior temporal areas. Hagoort and colleagues [2009] draw a distinction between integration and unification processes. Semantic integration occurs if different sources of information converge on a common memory representation. For example, the sight of a dog as well as the sound of its barking is both part of its representation in our knowledge about dogs [e.g., Hein et al., 2007]. Therefore, the posterior temporal lobe is mainly involved in the integration of highly overlearned, strongly associated material [Hein et al., 2007; Naumer et al., 2008; Willems et al., 2009], whereas the IFG integrates unrelated or incongruent AV combinations. In the latter case, semantic unification is always a constructive process in which a semantic representation is constructed that is not already available in memory. The temporal cortex is thought to contribute more to integration (perceptual matching or activation of a common memory representation), and the inferior frontal cortex is believed to play a stronger role in unification (see Hagoort et al. [2009]). This theoretical assumption has been applied to the integration of IC gestures [Willems et al., 2007, 2009], suggesting that

unification processes within the IFG are involved in the integration of IC coverbal gestures. However, in this study, activation of the IFG was predominantly associated with conditions in which gestures were mismatched with concrete sentence content. In terms of mismatches and the explicit presentation of conflicting information, frontal activation may also be explained by conflict processing [Kemotsu et al., 2005; Sætrevik and Specht, 2009], inhibition of a conflicting meaning [Hoenig and Scheef, 2009], or general top-down control or selection processes for appropriate information from the speech or gesture information stream [Badre et al., 2005; Gold et al., 2005; Moss et al., 2005; Thompson-Schill et al., 2005]. Based on findings from previous studies that used more natural control conditions than mismatch conditions (such as isolated speech, foreign language, or grooming gestures and speech), one may assume that integration processes occurring in posterior temporal areas are sufficient for the semantic integration of IC gestures in the context of concrete sentences [Green et al., 2009; Holle et al., 2008]. For the processing of MP coverbal gestures, in contrast, constructive unification processes within the IFG may be necessary [Kircher et al., 2009; Straube et al., 2009]. However, a direct comparison of the neural processing of natural IC and MP coverbal gestures has not yet been conducted.

In this study, we directly compared the processing of videos of natural IC and MP coverbal gestures. We hypothesized that the processing of IC coverbal gestures mainly relies on integration processes located in the posterior temporal lobe [Green et al., 2009; Holle et al., 2008], whereas the processing of MP gestures involves both integration processes located in posterior temporal brain regions and unification processes located in inferior frontal brain regions [Kircher et al., 2009; Straube et al., 2009]. To investigate bimodal processing without mismatch or other unnatural manipulations, we compared the bimodal speech and gesture items with unimodal control conditions. Previous investigations have already shown that the increased activation in posterior temporal and inferior frontal regions identified by such a comparison cannot be explained solely by bimodal processing, as speech semantics necessarily lead to increased activation (see Green et al. [2009] and Kircher et al. [2009]). Therefore, for this study, we believed that speech and gesture integration processes are reflected in additional activation in the bimodal conditions in contrast to unimodal speech and gesture conditions. These activation increases were calculated using conjunction analyses (see below). To ensure that these regions identified by conjunction analyses receive input from both modalities in isolation, we further restricted these analyses to regions that were also activated during the processing of the isolated conditions in contrast to baseline (fixation cross). To avoid processing related to “unnaturalness” or conflicts, we did not use mismatch manipulations in this study. However, there is already good evidence that the left posterior temporal lobe as well as the left IFG respond to mismatch manipulations [Book-

heimer, 2002; Friederici et al., 2003; Kuperberg et al., 2003] and unrelated speech and gesture information [e.g., Green et al., 2009; Willems et al., 2007, 2009]. One important component of IC and MP coverbal gestures is that the gestures in isolation are meaningless and have a particular meaning only in conjunction with speech [Feyereisen et al., 1988]. Thus, the alternative explanation of the increased activation of bimodal versus unimodal conditions as a double representation of two separate unimodal pieces of information (as recently demonstrated using simple unimodal and bimodal stimuli such as object names and pictures of objects [Hocking and Price, 2008]) is not likely to account for the activation pattern revealed by bimodal in contrast to unimodal speech and gesture items. This is because gestures in isolation have no particular semantic representation, unlike object names or object pictures. Therefore, we think that it is not necessary to include mismatch manipulations in this study to detect integration processes of speech and gesture information.

## METHODS

### Participants

Seventeen right-handed [Oldfield, 1971] healthy male volunteers participated in the study, all native German speakers (mean age = 27.8 years; range, 19–47 years) with no impairments of vision or hearing. None of the participants had any medical, neurological, or psychiatric illness, past or present. One subject had to be excluded from the analyses due to extensive movement (>5 mm) during fMRI-data acquisition. All participants gave written informed consent and were paid 20 Euro for participation. The study was approved by the local ethics committee.

### Stimulus Construction

The stimulus material consisted of video clips (each with a duration of 5 s) that presented an actor performing different combinations of speech and gestures: (1) IC coverbal gestures (concrete sentence content), (2) MP coverbal gestures (abstract sentence content) and two control conditions including (3) sentences without gestures (S; concrete sentence content), and (4) gestures without sentences (G, see Fig. 1). The gestures in the IC condition illustrate the form, size, or movement of something concrete that is mentioned in the accompanying speech [McNeill, 1992], whereas those in the MP condition illustrate the form, size, or movement of something abstract that is mentioned in the associated speech [McNeill, 1992]. The sentences in the control condition (S) are similar to the concrete sentences accompanied by IC gestures. Concrete sentences were used to control for general speech input processing.

A male actor was instructed to speak each sentence in conjunction with the associated gesture in the IC and MP

Co-verbal gesture conditions



Unimodal control conditions



Figure 1.

Examples of the four experimental conditions. The stimulus material consists of videos of an actor performing iconic (IC) and metaphoric (MP) coverbal gestures as well as two unimodal control conditions (spoken sentences without gestures [S] and gestures without sentences [G]). A screenshot of a typical video is

condition and without any arm or hand movement in the S condition. The G condition contains isolated gestures that were naturally associated with concrete sentence content, that is, IC. The sentences from the IG and MG conditions were also recorded without gestures for a subsequent memory task.

The actor stood with his hands at his sides before and after the speech and/or gesture. Each stimulus video clip was 5,000 ms in duration including 500 ms before and after the scene during which the actor neither spoke nor moved. This was done to account for the variable lengths of the speech and gestures and standardize the videos. Importantly, the execution of the gestures was actor-driven so as to obtain maximal naturalness (for a more detailed description of the stimulus material production and evaluation, see Kircher et al. [2009], for MP gesture integration, Straube et al. [2009]; for memory binding of speech and MP gestures, Green et al. [2009]; for IC and unrelated gesture integration, and Straube et al. [2010]; for a comparison of IC and emblematic coverbal gestures. In these studies,

shown for each condition. For illustrative purposes, the spoken German sentences were translated into English and written in speech bubbles. The IC and MP conditions differ in their sentence content (IC refers to concrete content whereas MP refers to abstract concepts).

parts of the stimulus material were used in event-related designs and different subject samples).

Twenty additional naïve German-speaking participants, who did not take part in the fMRI experiment, rated the stimulus videos on comprehension, naturalness, and imageability on a scale from 1 to 7 (1 = very low to 7 = very high). Rating results and further timing parameters for the four conditions are presented in Table I. Analyses of variance were performed for the rating parameters between the conditions. We found significant main effects for all variables [comprehension:  $F(3, 116) = 611.922, P < 0.001$ ; naturalness:  $F(3, 116) = 88.428, P < 0.001$ ; imageability:  $F(3, 116) = 148.802, P < 0.001$ ]. Post hoc analyses indicate a significant decrease in comprehension for the isolated gesture condition (G) in contrast to the speech conditions (IC, MP, and S; for all  $P < 0.001$ ). There were no significant differences in comprehension between the speech conditions (all  $P > 0.10$ ). Both combined conditions (IC and MP) were rated as significantly more natural than the unimodal conditions (all  $P < 0.001$ ).

TABLE I. Stimulus parameters

Parameter	Condition	Mean	SD	CI (95%)	
Comprehension	IC	6.8180	0.12308	6.7720	6.8640
	G	3.1013	0.73865	2.8255	3.3771
	S	6.5817	0.21301	6.5021	6.6612
	MP	6.7717	0.19813	6.6977	6.8456
Naturalness	IC	5.0780	0.48199	4.8980	5.2580
	G	3.6090	0.47952	3.4299	3.7881
	S	3.5760	0.33607	3.4505	3.7015
	MP	4.8750	0.54847	4.6702	5.0798
Imageability	IC	6.0700	0.37001	5.9318	6.2082
	G	3.7193	0.62537	3.4858	3.9528
	S	4.2797	0.21255	4.2003	4.3590
	MP	4.7017	0.48716	4.5198	4.8836
Speech duration	IC	2.3240	0.35658	2.1909	2.4571
	S	2.3583	0.25078	2.2647	2.4520
	MP	2.5440	0.27901	2.4398	2.6482
Gesture duration	IC	2.6353	0.43630	2.4724	2.7983
	G	2.8713	0.56973	2.6586	3.0841
	MP	2.3803	0.38825	2.2354	2.5253
Movement size	IC	2.1000	1.02889	1.7158	2.4842
	G	2.1333	0.93710	1.7834	2.4833
	MP	2.3000	0.70221	2.0378	2.5622

Rating results for the stimuli used in our experiment. Each condition consisted of 30 items, which were rated by 20 healthy subjects on a scale from 1 to 7 on “comprehension,” “imageability,” and “naturalness” (1 = very low to 7 = very high). Speech duration (measured from speech onset to speech offset), gesture duration (measured from arm movement onset to arm movement offset), and movement size (number of rectangles crossed in the gesture space) are listed as well. The above table shows mean, standard distribution (SD), and 95% confidence interval (CI).

There were no differences in naturalness within the bimodal (IC vs. MP,  $P < 0.20$ ) and unimodal conditions (S vs. G,  $P < 0.20$ ). Imageability ratings indicated that items in the IC condition elicited mental images for the participants, but less so for items in the MP, G, and S conditions (all  $P < 0.005$ ). These low values for imageability in the MP condition reflect the high degree of abstractness.

The average duration of concrete sentences (measured from speech onset to speech offset) did not vary significantly [IC vs. S:  $t(58) = 0.431$ ,  $P = 0.668$ ; see Table I]. Sentences in the MP condition were slightly longer than those in the IC and S conditions [MP > IC:  $t(58) = 2.661$ ,  $P < 0.05$ ; MG > S:  $t(58) = 2.711$ ,  $P < 0.01$ ; see Table I] and had slightly shorter gesture durations (average difference = 0.25 s, measured from arm movement onset to arm movement offset) than the IC condition [IC < MP:  $t(58) = 2.391$ ,  $P < 0.05$ ]. Despite being statistically significant, these differences in speech and gesture duration are small (average difference = 0.25 s) and due to the blocked presentation (see below) they should not influence the analyses and claims of our study. Furthermore, memory performance at the single item level for coverbal gesture stimuli is not correlated with speech duration ( $r = -0.037$ ,  $P < 0.845$ ), gesture duration ( $r = 0.039$ ,  $P < 0.837$ ), and the three rating parameters (naturalness:  $r = 0.117$ ,  $P < 0.537$ ; comprehension:  $r = 0.003$ ,  $P < 0.989$ ; imageability:  $r = -0.164$ ,  $P < 0.386$ ). These behavioral data indicate that differences in speech and gesture dura-

tion have little to no influence on speech and gesture encoding. Furthermore, there are no significant differences in hand use (IC: 17 right/13 both hands; G: 15 right/15 both hands; MP: 20 right/10 both hands; Chi-square = 1.731,  $df = 2$ ,  $P = 0.421$ ) and movement size between conditions including gestures (IC, G, and MP; see Table I). To compare the movement size between conditions, we coded each video clip with regard to the extent of the hand movement. We divided the video screen into small rectangles that corresponded to the gesture space described by McNeill [1992, 2005] and counted the number of rectangles in which gesture movements occurred. We found no differences between conditions [ $F(2, 87) = 0.425$ ,  $P = 0.655$ ; see Table I].

## Experimental Design

Thirty stimuli from each of the four experimental conditions were presented in a block design. Each block consisted of five videos of the same condition and was 25 s in length (five videos  $\times$  5 s). In total, six blocks of each condition were presented in a pseudorandomized order and separated by a baseline condition (gray background with a fixation cross) with a duration of 15 s, during which the fixation cross shortly disappeared two times (about every 5 s). Each participant saw 120 video clips during the functional measurement, which lasted a total of 17 min.

Subjects were instructed to watch the videos and to respond each time they saw a new picture appear (either the video or baseline fixation cross) by pressing a button with the left index finger. This was done to ensure that they paid attention during all conditions and baseline. This implicit-encoding task was chosen to focus participants' attention on the middle of the screen and enabled us to investigate implicit speech and gesture processing. Before the scanning session, each participant participated in at least 10 practice trials outside the scanner that were different from those used in the experiment. Additional clips were presented during the overview scans, and the volume was adjusted, so that the sentences could be well heard.

### Subsequent Memory Task

A subsequent memory task was conducted to investigate the effects of speech and gesture integration on behavior. It was hypothesized that successful integration of speech and gesture would lead to increased memory performance in the subsequent memory task [Straube et al., 2009]. Therefore, recognition memory performances for the three conditions (IC, MP, and S) were obtained around 10 min after scanning. Subjects were shown only videos without gestures to create similar recognition circumstances for all conditions. Fifteen spoken sentences from each of the IC, MP, and S conditions were presented with an equal number of new sentences unaccompanied by gestures for each condition. Thus, 30 new sentences with concrete content and 15 new sentences with abstract content were intermixed with the old sentences during the recognition task. In total, 90 videos of spoken sentences without gestures were presented in the recognition phase. Participants indicated with an "old"/"new" response if the actor had spoken the sentence during the scanning session, regardless if there had been a gesture or not ("old," left button; "new," right button).

Presentation software (Version 9.2, 2005) was used for stimulus presentation and response measurement in both the fMRI and the recognition experiment.

### fMRI Data Acquisition

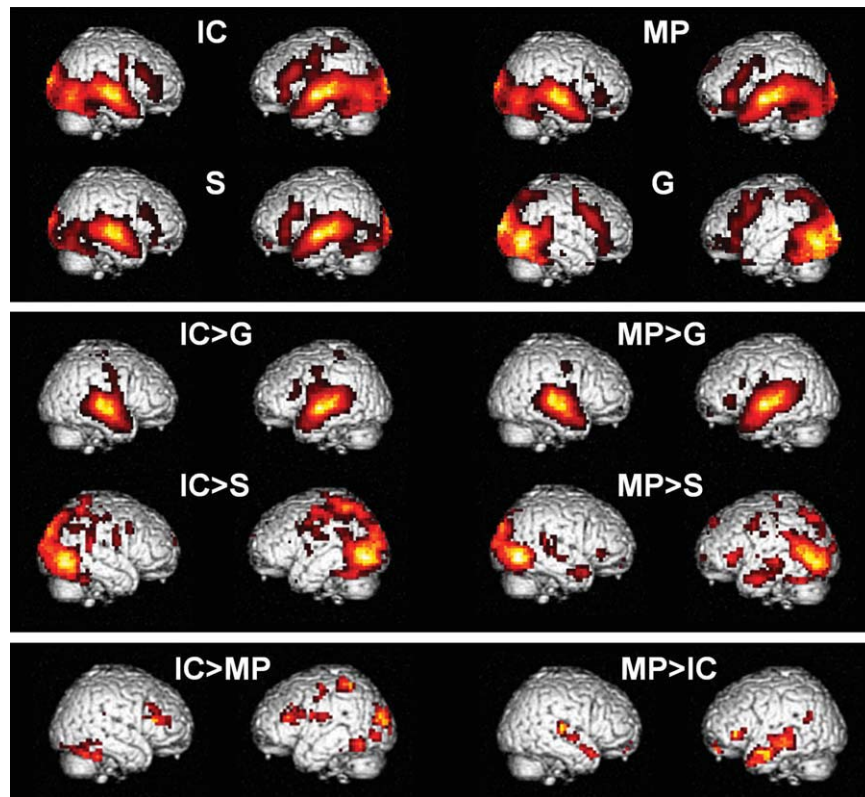
MRI was performed on a 1.5T Philips scanner (Philips MRT Achieva series). Functional data were acquired with echo planar images in 31 transversal slices [repetition time (TR) = 2800 ms; echo time (TE) = 50 ms; flip angle = 90; slice thickness = 3.5 mm; interslice gap = 0.35 mm; field of view (FoV) = 240 mm, voxel resolution = 3.75 × 3.75 mm]. Slices were positioned to achieve whole brain coverage. Three hundred and sixty volumes were acquired during the functional run. After the functional run, an anatomical scan was acquired for each participant using a high-resolution T1-weighted 3D-sequence consisting of 140 transversal slices (TR = 7,670 ms; TE = 3,500 ms; FoV = 256 mm; slice thickness = 1 mm; interslice gap = 1 mm).

### Data Analysis

MR images were analyzed using Statistical Parametric Mapping (SPM; [www.fil.ion.ucl.ac.uk](http://www.fil.ion.ucl.ac.uk)) implemented in MATLAB 6.5 (Mathworks, Sherborn, MA). For practical reasons, minimally error-prone script-based preprocessing and first-level analyses were performed with SPM2. The resultant outputs are comparable to and compatible with the newer SPM5-version that was used for the second-level analyses. The first five volumes of each functional run were discarded from the analysis to minimize T1-saturation effects. To correct for different acquisition times, the signal measured in each slice was shifted relative to the acquisition time of the middle slice using a slice interpolation in time. All images of one session were realigned to the first image of a run to correct for head movement and normalized into standard stereotactic anatomical MNI-space by using the transformation matrix calculated from the first EPI-scan of each subject and the EPI-template. Afterward, the normalized data with a resliced voxel size of 4 × 4 × 4 mm were smoothed with a 10-mm Full Width at Half Maximum (FWHM) isotropic Gaussian kernel to accommodate for intersubject variation in brain anatomy. A high-pass filter (128 s cut of period) was applied to remove low-frequency fluctuations in the BOLD signal.

The expected hemodynamic response at the onset of each block was modeled with a duration of 25 s by two response functions, a canonical hemodynamic response function [HRF; Friston et al., 1998] and its temporal derivative. The temporal derivative was included in the model to account for the residual variance resulting from small temporal differences in the onset of the hemodynamic response, which is not explained by the canonical HRF alone. The functions were convolved with the block sequence in a general linear model.

To correct the fMRI results for multiple comparisons, we used a Monte-Carlo simulation of the brain volume to establish an appropriate voxel contiguity threshold ([Slotnick et al., 2003]; also used by Slotnick and Schacter [2004] and Garoff et al. [2005]). This correction is more sensitive to smaller effect sizes while still correcting for multiple comparisons across the whole brain volume. The procedure is based on the fact that the probability of observing clusters of activity due to voxel-wise Type I error (i.e., noise) decreases systematically as cluster size increases. Therefore, the cluster extent threshold can be determined to ensure an acceptable level of corrected cluster-wise Type I error. We ran a Monte-Carlo simulation to model the brain volume ([http://www2.bc.edu/\\_slotnics/scripts.htm](http://www2.bc.edu/_slotnics/scripts.htm); with 1,000 iterations) using the same parameters as in our study (i.e., acquisition matrix, number of slices, voxel size, resampled voxel size, and FWHM). Assuming an individual voxel type I error of  $P < 0.05$ , a cluster extent of 17 contiguous resampled voxels was indicated as necessary to correct for multiple-voxel comparisons at  $P < 0.05$ . We applied this cluster threshold to all reported analyses.



**Figure 2.**

The activation pattern for each of the four conditions in contrast to baseline (fixation cross; top), the bimodal conditions in contrast to the unimodal conditions (IC > S; IC > G; MP > S; MP > G; middle) and the direct contrasts of the processing of iconic (IC) and metaphoric (MP) coverbal gestures (IC > MP, MP > IC; bottom). IC, iconic coverbal gesture condition (concrete); MP, metaphoric coverbal gestures condition (abstract); S, speech without gesture; G, gesture without speech.

The reported voxel coordinates of activation peaks are located in MNI space. The functional data were referenced to probabilistic cytoarchitectonic maps for the anatomical localization [Eickhoff et al., 2005].

Statistical analyses of data other than fMRI were performed using SPSS version 14.0 for Windows (SPSS, Chicago, IL). Discrimination performance ( $d'$ ) and response criterion ( $c$ ) were calculated following the signal detection theory ( $d' = z(\text{Hits}) - z(\text{FA})$ ;  $c = -[1/2] [z(\text{Hits}) + z(\text{FA})]$ ; e.g., Macmillan and Creelman [1991]). Statistical analyses are two-tailed with  $\alpha$ -levels of significance of  $P < 0.05$ .

### Contrasts of Interest

We applied particular contrasts of interest to test our hypotheses. We analyzed and preset activation patterns of all four conditions against baseline to obtain an overview of the general brain regions involved in isolated and combined speech and gesture processing (baseline contrasts: IC, S, G, and MP; see Fig. 2).

To test our hypothesis that there would be differences in the processing of IC versus MP coverbal gestures, we calculated difference contrasts between both conditions (IC vs. MP and vice versa; see Fig. 2 and Table III).

In the context of this study, we define integration as the increased neural activity due to bimodal (as opposed to unimodal) processing of speech and gesture. To identify bimodal integration activity, we calculated the conjunction (testing a logical AND, Nichols et al. [2005]) of bimodal in contrast to both unimodal conditions (e.g.,  $[(IC > S \cap IC > G) \cap S \cap G]$ ). To confirm that the corresponding regions do in fact receive input from both unimodal conditions (see Calvert et al. [2001]), we further restricted the analysis to regions activated during both unimodal conditions when compared with baseline. Therefore, we calculated the following contrast for the IC coverbal gesture condition:  $[(IC > S \cap IC > G) \cap S \cap G]$  and the following conjunction for MP coverbal gesture condition:  $[(MP > S \cap MP > G) \cap S \cap G]$ .

Finally, we were interested in the similarities and differences between the multimodal integration processes of IC and MP coverbal gestures. Therefore, we calculated the

conjunction  $[(IC > S \cap IC > G) \cap (MP > S \cap MP > G) \cap IC \cap MP \cap S \cap G]$  and the contrast between coverbal gesture types within bimodal processing regions:  $[(MP > IC) \cap (MP > S \cap MP > G) \cap IC \cap S \cap G]$  and  $[(IC > MP) \cap (IC > S \cap IC > G) \cap MP \cap S \cap G]$ .

## RESULTS

### Behavioral Results

Behavioral data from six subjects in the fMRI-attention task (button press after each video) and three from the subsequent recognition task are missing due to technical reasons. The remaining data sets reveal continuous responses over the whole session with no differences in reaction times across conditions (IC  $m = 530$  ms,  $SD = 247$  ms; MP  $m = 555$  ms,  $SD = 272$  ms; G  $m = 513$  ms,  $SD = 188$  ms; S  $m = 507$  ms,  $SD = 238$  ms;  $F(3, 8) = 0.318$ ,  $P = 0.812$ , within-subject ANOVA).

### Subsequent Memory Task

The subsequent memory task was conducted to obtain behavioral evidence of successful integration processes during encoding in the bimodal conditions. Only videos of spoken sentences without gestures were presented in the subsequent memory task, so that the recognizable characteristics were the same between conditions (IC, S, and MP). Therefore, the labels "IC gesture condition" and "MP gesture condition" in the following analyses refer to the encoding phase in which concrete sentence content was accompanied by IC gestures and abstract sentence content by MP gestures (MP). The proportion of sentences correctly endorsed as old (hits) was 54% ( $SD = 17\%$ ) for the IC gesture condition (concrete speech content), 58% ( $SD = 23\%$ ) for the MP gesture condition (abstract sentence content), and 43% ( $SD = 15\%$ ) for the no-gesture condition (S; concrete sentence content), with a false alarm rate (FA) of 15% for spoken sentences with concrete sentence content and 28% for spoken sentences with abstract sentence content. The hit rate is significantly higher than the FAs rate in all conditions (hits  $>$  FA), indicating that successful encoding of memories occurred during fMRI-data acquisition [IC:  $t(12) = 7.761$ ,  $P < 0.001$ ; MP:  $t(12) = 5.078$ ,  $P < 0.001$ ; S:  $t(12) = 7.840$ ,  $P < 0.001$ ]. Pairwise comparisons for hits and FA indicated that the hit rates were significantly increased in the gesture conditions (IC and MP) in comparison with the no gesture condition [IC  $>$  S:  $t(12) = 2.609$ ,  $P < 0.05$ ; MP  $>$  S:  $t(12) = 3.161$ ,  $P < 0.01$ ] and that the FA rate was significantly reduced for concrete in comparison with abstract sentence content [ $t(12) = -3.456$ ,  $P < 0.01$ ]. The hit rate between the gesture conditions did not differ [IC  $>$  MP:  $t(12) = -0.737$ ,  $P = 0.475$ ].

The analysis of hits and FA revealed differences in the response behaviors of the participants across the conditions. Therefore, we calculated the discrimination performance ( $d'$ ) between old and new items independent of the individual response criteria ( $c$ ) by conducting a signal

detection analysis. FA rate of concrete sentences was used for both IC and S, because there was no difference between the concrete spoken sentences other than if they were presented with or without gestures (as in our recognition experiment). We found that the discrimination performance ( $d'$ ) for the IC condition was not significantly different from the one for MP [ $t(12) = -1.872$ ,  $P = 0.086$ ] but was greater in comparison with the S condition [ $t(12) = 2.538$ ,  $P < 0.05$ ]. In contrast, response criteria ( $c$ ) were significantly enhanced in IC in contrast to the MP [ $t(12) = -2.928$ ,  $P < 0.05$ ] and S condition [ $t(12) = -2.534$ ,  $P < 0.05$ ]. The discrimination performance ( $d'$ ) for the MP condition was not significantly different from the S condition [ $t(12) = -0.505$ ,  $P = 0.622$ ].

The data from the memory task revealed increased hit rates for sentences previously accompanied by gesture in comparison with sentences previously presented in isolation. Because of the increased FA rate for the abstract sentences, we found no significant difference in discrimination performance ( $d'$ ) between the MP condition and the S and IC condition. However, we found an enhanced memory performance for the IC condition, which was also indicated by the discrimination parameter ( $d'$ ). These data suggest that, during encoding, gestures enhance the elaboration or foundation of a memory representation, thus leading to enhanced performance during a recognition task, regardless of whether gestures were present in the test situation.

### fMRI Results

We first show the results of the direct contrasts of each condition against baseline (fixation cross). Second, we directly compare the processing of IC and MP gestures. Finally, the conjunction analyses designed to reveal activations related to natural semantic integration processes are gradually developed and finally checked for overlaps and differences between IC and MP speech-gesture processing.

### Conditions in Comparison with Low-Level Baseline (Fixation Cross)

For all auditory conditions (IC, S, and MP) versus baseline (fixation cross), we found an extended network of bihemispheric medial and lateral temporal and left inferior frontal activity. In the gesture conditions (IC, G, and MP), activation clusters extended into the occipital lobes and the cerebellum (see Fig. 2). Gesture conditions without speech show less activation in the temporal lobes than do conditions with auditory input (see Fig. 2).

### IC Versus MP Coverbal Gestures

The direct comparison of iconic (IC) versus metaphoric (MP) speech-gesture pairs (IC  $>$  MP) resulted in large activation clusters in inferior temporal-occipital and inferior frontal (BA 45) regions bilaterally, in the left superior



**TABLE II. Iconic versus metaphoric coverbal gestures**

Anatomical region	Cluster extent	Coordinates			No. of voxels	t-value
		x	y	z		
MP > IC						
Left middle temporal gyrus	Left ITG, temporal pole	-56	4	-24	263	4.07
Left superior orbital gyrus	Right superior orbital gyrus	-8	52	-20	36	3.87
Right thalamus	Right, amygdala, HC	16	-24	12	22	3.61
Left inferior frontal gyrus	Left BA 45	-56	28	0	41	3.00
Right superior temporal gyrus		64	-24	8	76	2.73
Right medial temporal pole	Right temporal pole	60	4	-16	41	2.47
Left middle temporal gyrus		-44	-56	20	17	2.46
IC > MP						
Left inferior parietal lobe	Left SPL; BA 2, 3, 4, 6	-32	-48	60	68	4.85
Right middle cingulate cortex	White matter	20	0	32	36	4.80
Right inferior frontal gyrus	Right BA 45, MFG	52	32	16	64	3.98
Right inferior parietal cortex	Right insula	36	-36	28	44	3.76
Right fusiform gyrus	Right ITG, cerebellum	36	-36	-16	135	3.40
Left fusiform gyrus	Left pallidum	-32	-36	-16	95	3.34
Left postcentral gyrus	Left BA 3, 6, 44, Rol. OP	-60	0	20	44	3.32
Left inferior temporal gyrus	Left BA 3, insula, Rol.OP	-52	-60	-12	51	3.32
Right amygdala	Right parahippocampus	24	0	-12	56	3.26
Left inferior frontal gyrus	Left BA 44, 45	-40	20	12	73	3.17
Left precentral gyrus	Left BA 6; postcentral gyrus	-36	-12	56	30	3.15
Left superior occipital gyrus	Left MOG, SOG, BA 17, 18	-16	-92	12	149	3.05
Left inferior occipital gyrus		-44	-80	-4	32	2.94
Right lingual gyrus	Right BA 18; HC	12	-40	0	26	2.92

Coordinates (MNI), cluster extent, and t-values of the difference contrasts of iconic versus metaphoric gestures. Cluster level corrected at  $P < 0.05$ . HC, hippocampus; IFG, inferior frontal gyrus; ITG, inferior temporal gyrus; MFG, medial frontal gyrus; MTG, middle temporal gyrus; Rol.OP, rolandic operculum; SPL, superior parietal lobe; STG, superior frontal gyrus.

parietal lobule, and in the left occipital lobe (see Table II and Fig. 2). In contrast, the reverse comparison (MP > IC) resulted in more anterior activations of the bilateral middle and superior temporal gyri, the left ventral part of the IFG (BA 45) and the left superior orbital gyrus (see Table II and Fig. 2 for the complete activation pattern).

### Multimodal Integration of IC and MP Coverbal Gestures

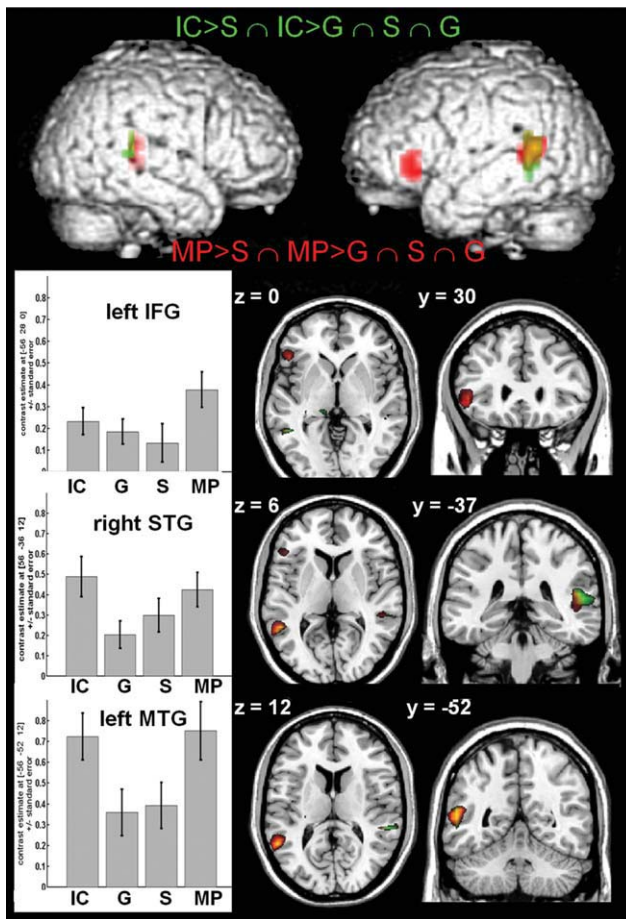
For bimodal integration processes in the iconic (IC) coverbal gesture condition (conjunction analysis [(IC > S ∩ IC

> G) ∩ S ∩ G]), we observed activation of the left posterior temporal lobe (MTG/STG) and the right superior temporal gyrus (STG; see Table III and Fig. 3). The corresponding analyses for metaphoric (MP) coverbal gestures [(MP > S ∩ MP > G) ∩ S ∩ G] revealed activity in left posterior temporal lobe (MTG/STG), the right superior temporal gyrus (STG) and the left IFG (see Table III and Fig. 3). The common activation analysis for the integration of both coverbal gesture types (IC > S ∩ IC > G ∩ MP > S ∩ MP > G ∩ S ∩ G ∩ MP ∩ IC) revealed activation in the left posterior middle temporal gyrus (MTG; MNI  $x, y, z$ : -56, -52, 12;  $t = 2.9$ ; 27 voxels). With regard to the differences in multimodal integration processes between IC and MP coverbal

**TABLE III. Integration of speech and gesture**

	IC-integration			MC-integration			MP > IC-integration		
	$x,y,z$ (mm)	Cluster	t-value	$x,y,z$ (mm)	Cluster	t-value	$x,y,z$ (mm)	Cluster	t-value
Left MTG	-56 -52 12	32	2.90	-56 -52 12	52	3.78	—	—	ns
Left IFG	—	—	ns	-52 28 -4	32	2.95	-52 28 0	19	2.46
Right STG	56 -36 12	17	2.87	48 -36 16	19	2.62	—	—	ns

Coordinates (MNI), cluster size (no. of voxels) and t-values of the whole-brain analyses cluster level corrected at  $P < 0.05$  for the integration of iconic and metaphoric gestures. IC-integration, conjunction of [(IC > S) ∩ (IC > G) ∩ S ∩ G]; MP-integration, conjunction of [(MP > S) ∩ (MP > G) ∩ S ∩ G]; MP > IC-integration, conjunction of [(MP > IC) ∩ (MP > S) ∩ (MP > G) ∩ S ∩ G ∩ IC]; IFG, inferior frontal gyrus; MTG, middle temporal gyrus; STG, superior frontal gyrus.



**Figure 3.**

Regions with increased activation for bimodal (IC, green; MP, red; IC&MP, yellow) in contrast to unimodal processing of speech and gesture, as identified through conjunction analyses. Green:  $[(IC > S \cap IC > G) \cap S \cap G]$ , Red:  $[(MP > S \cap MP > G) \cap S \cap G]$ , Yellow:  $[(IC > S \cap IC > G) \cap (MP > S \cap MP > G) \cap S \cap G]$ . Only the activation of the left IFG (19 voxels; see Table III) is also significant for the following conjunction analysis:  $[(MP > IC) \cap (MP > S \cap MP > G) \cap IC \cap S \cap G]$ . IC, iconic coverbal gesture condition (concrete); MP, metaphoric coverbal gestures condition (abstract); S, speech without gesture; G, gesture without speech.

gestures, we observed more activation during the MP in contrast to the IC condition ( $MP > IC \cap MP > S \cap MP > G \cap IC \cap S \cap G$ ) within the left IFG (see Table III and Fig. 3). No significant activation was observed for the reversed contrast direction ( $IC > MP \cap IC > S \cap IC > G \cap IC \cap S \cap G$ ).

These data reveal the different neural mechanisms involved in the processing of IC and MP gestures. We observed increased activation in the left posterior middle temporal gyrus for both IC and MP coverbal gestures due to bimodal speech-gesture presentation. The left IFG is

additionally activated only for MP gestures. Both of these regions are also activated during the unimodal conditions in contrast to baseline, supporting the idea that these areas also receive input from auditory and visual modalities in isolation and thus can be regarded as sites of neural integration. To determine if differences in response behavior (c) between IC and MP during the subsequent memory task might be a confounding variable with regard to the integration processes, we included c as covariate in the group analyses. However, all reported activation patterns remained the same (with only minor changes in cluster size and t-values) after inclusion of the covariate. Thus, differences in activation between IC and MP conditions cannot be explained by differences in response behavior and appear to reflect differing levels of abstractness between speech and gesture pairings.

## DISCUSSION

Gestures are an important component of interpersonal communication (see McNeill [1992, 2005]) and have a huge impact on speech perception [Holle et al., 2008], memory [Straube et al., 2009], and their underlying brain processes [Dick et al., 2009; Green et al., 2009; Holle et al., 2008; Hubbard et al., 2009; Kircher et al., 2009; Straube et al., 2009; Willems et al., 2007, 2009]. However, past investigations of speech and gesture integration have not directly compared the processing of different coverbal gestures. This study reveals the shared and unique bimodal processes that are involved in the integration of IC and MP coverbal gestures and also demonstrates the specific functional roles of posterior temporal and inferior frontal brain regions.

In line with our hypotheses, IC coverbal gestures were associated with increased activity in the posterior temporal lobe (BA 37) in contrast to the isolated control conditions. This region was also active during conditions of speech and gesture in isolation as well as for MP coverbal gestures. Therefore, we consider it as the area responsible for the integration of speech and gesture information. This assumption is supported by past research on speech and gesture integration [Dick et al., 2009; Green et al., 2009; Holle et al., 2008; Kircher et al., 2009]. We found that this same region as well as the left IFG was activated for MP gestures in contrast to the isolated control conditions. This activation of the IFG was also greater than that of IC gesture conditions. This finding is consistent with past observations of posterior temporal and inferior frontal activation for MP gesture processing [Kircher et al., 2009; Straube et al., 2009] and the lack of evidence for frontal activations for the integration of IC coverbal gestures [Green et al., 2009; Holle et al., 2008]. IC and MP coverbal gesture processing were both associated with the additional involvement of regions of the right homologue of the left posterior temporal activations, a pattern of activation that is often observed in language studies [Hagoort

et al., 2009], but may also suggest that the right hemisphere aids in the integration of speech and gesture information [Kircher et al., 2009]. Our data provide evidence of common integration processes in posterior temporal areas for IC and MP coverbal gestures and suggests that the left IFG plays a unique role in the integration of MP gestures.

Several studies report that the posterior temporal lobe (STS/MTG) is an important multimodal integration site [Amedi et al. 2005; Beauchamp et al., 2004a,b; Callan et al., 2004; Calvert, 2001; Calvert and Thesen, 2004; Hein and Knight, 2008]. However, the IFG has also been proposed to play a role in multimodal integration when the congruency and novelty of picture and sound are modulated [Hein et al., 2007; Naumer et al., 2008]. We found that the posterior temporal lobes are also involved in the integration of both IC and MP gestures and their corresponding sentence contexts. This finding is remarkable, because the connection between speech and gesture, especially at the abstract level of MP coverbal gestures, is quite different from the connection between simple sounds and corresponding images (e.g., the sound of barking and the sight of a dog). However, even at this abstract linguistic level, some multimodal components seem to be integrated within the posterior temporal lobe. Although both gesture types produced integration effects in temporal brain regions, the left inferior frontal lobe was only activated for MP gestures. Because abstractness rather than novelty or congruency was manipulated in our study, the differences in activation within the IFG between conditions appear to be due to the abstract relationships between concrete visual and abstract verbal information. Additional online integration or unification processes within the IFG appear to be involved in constructing such relationships.

These results and interpretations are in agreement with a recent claim about the separation of semantic integration and semantic unification processes in semantic comprehension [Hagoort et al., 2009], which suggests that both components are specifically involved in the processing of IC and MP coverbal gestures. Also in line with our pattern of results, Hagoort et al. [2009] associated integration with left posterior temporal areas and unification processes with inferior frontal brain regions. Our data suggest that integration processes (the activation of an already available common memory representation) appear to be all that is necessary for the processing of IC gestures in the context of concrete sentences. In contrast, unification processes (a constructive process in which a semantic representation that is not already available in memory is constructed) are necessary for the processing of MP coverbal gestures. The fact that IC and MP coverbal gesture processing do not differ in their posterior temporal activation patterns agrees with the assumption that a dynamic relationship between left IFC and left superior/middle temporal cortex is necessary for successful semantic unification when the integration load is high [Hagoort et al., 2009]. Therefore, our data suggest that IC coverbal gestures seem to activate directly the common memory representation of concrete speech

and gesture information (e.g., “round bowl” and the corresponding round gesture). In contrast, constructive processes (unification processes) within the left IFG appear to be necessary for the integration of gesture information (e.g., depicting an arch with the right hand) in the context of abstract sentences (e.g., “The politician build a bridge to the next topic”). This finding is consistent with studies that revealed posterior temporal but no inferior frontal activity during the processing of IC gestures [in contrast to grooming gestures: Holle et al., 2008; in contrast to gestures in context of a foreign language: Green et al. [2009] and with findings that both posterior temporal and inferior frontal regions are involved in the processing of MP coverbal gestures [Kircher et al., 2009; Straube et al., 2009]. However, our result conflicts with the theory that unification processes within the left inferior frontal cortex are necessary for the comprehension of IC coverbal gestures [Willems et al., 2007, 2009]. This claim is predominantly based on the finding that the left inferior frontal lobe is sensitive to mismatch manipulations of speech and gesture [Willems et al., 2007]. However, in such a paradigm, frontal activation may be explained by conflict processing [Kemmons et al., 2005; Sætrevik and Specht, 2009], inhibition of a conflicting meaning [Hoenig and Scheef, 2009], general top-down control [January et al., 2009], or selection processes for appropriate information from the speech and gesture information stream [Badre et al., 2005; Gold et al., 2005; Moss et al., 2005; Thompson-Schill et al., 2005]. A recent study of our working group replicated the finding that the inferior frontal region is involved in the processing of gestures that are incongruent with speech but is not involved in the processing of speech-congruent gestures [Green et al., 2009]. This suggests that the role of the IFG in IC speech-gesture processing is not purely integrative but rather is related to the detection and resolution of incompatible stimulus representations and reanalyzing misinterpretations [Kuperberg et al., 2008; Novick et al., 2005]. In this study, we revealed distinct multimodal semantic integration processes within the inferior frontal lobe for the processing of MP, but not IC, coverbal gestures. Similar to Willems et al. [2007], we found a significant increase in activity in the IFG for IC coverbal gestures in contrast to low-level baseline (fixation cross “+”). However, we also observed the same activation during the unimodal conditions, which does not suggest that there is significant multimodal processing in the IC condition. The focus of our study is on multimodal processes in which the multimodal condition (IC or MP) is associated with more activity than the unimodal conditions (S and G). These multimodal processes appear to be located in bilateral posterior temporal regions for both IC and MP coverbal gestures as well as in the IFG for MP coverbal gestures. Contrary to our study, Willems et al. [2009] showed that in comparison with unimodal conditions, IC coverbal gestures are associated with increased activation of the IFG. Through a direct comparison of IC and MP coverbal gestures we also found increased bilateral

activation in a more posterior–dorsal part of the IFG (BA 44/45; which corresponds to the region identified by Willems et al. [2009]). However, in our study, this region was not significantly activated during conditions of speech and gesture in isolation. This discrepancy may be due to differences in stimulus material (for example, Willems et al. [2009] repeated each video three times, which may have led to reduced variance between conditions and more than one gesture could occur for a single item). However, we observed increased activation in a more anterior–ventral part of the IFG (BA45/47) for MP gestures in contrast to isolated conditions. Therefore, the IFG's contribution to the processing of IC and MP gestures (whether multimodal or not) may be further distinguished by a dorsal–ventral gradient in which semantically abstract concepts are processed in more ventral regions in the inferior frontal lobe (see Shalom and Poeppel [2008]). Furthermore, in line with the recent differentiation of frontal lobe functions with regard to an anterior–posterior gradient of an abstract-representational hierarchy [Badre, 2008; Badre and Esposito, 2009], IC gestures activated a more posterior region than MP coverbal gestures, which also supports the anterior–posterior difference with regard to abstractness in our study.

Unification of speech and gesture semantics within the inferior frontal lobe appears to be important for the integration of MP coverbal gestures and may be interpreted as a process responsible for making inferences [Bunge et al., 2009], relational reasoning [Wendelken et al., 2008], and the building of analogies [Bunge et al., 2005; Green et al., 2006; Luo et al., 2003]. Those processes may also be involved in the comprehension of MP or ambiguous communications and consistently activate the left IFG [Ahrens et al., 2007; Eviatar and Just, 2006; Lee and Dapretto, 2006; Rapp et al., 2004, 2007; Stringaris et al., 2007]. In contrast, integration within the posterior temporal lobe appears to rely more on perceptual matching processes, such as those that are engaged in the identification of tool sounds [Beauchamp et al., 2004a,b; Lewis, 2006; Lewis et al., 2005; for a review], action words [Kable and Chatterjee, 2006; Kable et al., 2005] or highly familiar and overlearned stimulus associations [Hein et al., 2007; Naumer et al., 2008]. These representations may be readily available in memory [Hagoort et al., 2009]. Although unification processes were involved in the processing of MP coverbal gestures [Kircher et al., 2009; Straube et al., 2009] and speech-gesture mismatches [Green et al., 2009; Willems et al., 2007], integration processes are involved in the processing of all types of coverbal gestures (e.g., IC, MP, and unrelated) and even appear to be sufficient for the comprehension of natural IC coverbal gestures [Green et al., 2009; Holle et al., 2008]. This interpretation also corresponds with functional anatomic models of language ([see Shalom and Poeppel [2008]), suggesting that specific types of processing occur in particular cortical areas, for example, memorizing in the temporal cortex, analyzing in the parietal cortex, and synthesizing in the frontal cortex. Although

temporal structures are responsible for information stored in memory (e.g., words or names), the frontal lobe is involved in synthesizing such information. Shalom and Poeppel [2008] suggest that computations within the frontal lobe are likely to be abstract and somewhat generic. These synthesizing processes may be particularly important when speech and gesture are related but at an abstract level, as they are in MP coverbal gestures.

Although our findings support the anatomical separation of integration and unification processes proposed in the semantic unification theory [Hagoort et al., 2009], our data also suggest a different integration process than that proposed by Willems and colleagues [2007, 2009] for IC coverbal gestures. Our data imply that processes in the posterior temporal lobe are involved in the integration of natural IC coverbal gestures. Our study benefits from the absence of mismatch conditions that could easily lead to confounding processes within the frontal lobe. We demonstrated that natural gestures paired with concrete sentence content could be integrated within the posterior temporal lobe. In contrast, natural-related gestures paired with abstract sentence content led to additional activation of the frontal lobe. Behavioral data from the subsequent memory task support the theory that integration has relevant behavioral effects, as shown in the increased hit rate for IC and MP coverbal gestures compared to isolated speech conditions. Unlike previous investigations (Straube et al. 2009) we used a unimodal recognition task to assess the memory performance of our participants. In this task, only videos of spoken sentences without gestures were presented. Thus, our results demonstrate the visual influence of gestures during encoding, resulting in differences in storage processes for concrete and abstract sentence content. This design is advantageous in that the recognition condition is equal across all conditions. In spite of the similarity, we found that the gesture conditions were better remembered in the memory task. These effects most likely are due to the binding processes of speech and gesture information (see Straube et al. [2009]). Therefore, the lack of frontal activation during IC coverbal gesture conditions is not likely to be due to the absence of integration processes, but rather it suggests that the activation in posterior temporal lobe is necessary for successful integration leading to enhanced memory performance.

A limitation of this study is that the unimodal control condition for isolated speech consists of concrete sentences (similar to the IC coverbal gesture condition) and not abstract sentences. Concrete sentences were chosen to control for general speech processing. The problem with regard to the interpretation of our results might be that the activation increase found for MP gestures in the left IFG in contrast to the isolated control conditions is related to a simple difference in the abstractness between the control speech and MP gesture condition. However, in a previous study, we observed increased activation in the left IFG for MP coverbal gestures in contrast to control sentences with abstract semantic content [Kircher et al., 2009].

Thus, the finding of increased activation in the left IFG in this study is unlikely to be simply an artifact of the concrete speech control condition. Furthermore, the absence of IFG activation for IC gestures in contrast to the isolated control conditions cannot be explained by the choice of control condition, because similar concrete sentences were used for bimodal and unimodal conditions.

Gestures are an important part of interpersonal communication. They may illustrate physical properties (e.g., “the ball is round”) or abstract concepts and ideas (e.g., “a mental connection”). This study compares for the first time different kinds of illustrative and elaborative gestures that differ in their relation to speech. We found that there are both shared and unique mechanisms for the integration of IC and MP gesture and speech pairs. Although our data suggest that posterior temporal areas are involved in the integration of both gesture types, the left IFG appears to be specifically related to the processing of MP coverbal gestures. This supports the theory that integration processes differ in terms of perceptual-matching and higher-order relational processes and agrees with recent anatomical language models and the semantic unification theory. However, we still do not have a clear understanding of the specific contribution of both integration sites and integration levels to the creation of a common representation and its functional relevance. Nonetheless, the ability to distinguish between different types of speech and gesture pairs appears to be important, as the integration processes for each seem to be different. In addition, it is important to have natural speech and gesture pairs, so as to avoid confounding processes that activate frontal areas, such as mismatch detection or conflict processing due to low naturalness.

## ACKNOWLEDGMENTS

The data were collected in Aachen and the data analyses and the writing of the manuscript primarily occurred in Philadelphia (US; B.S., B.B.) and in Marburg (Germany; T.K., A.G., B.S.).

## REFERENCES

- Ahrens K, Liu HL, Lee CY, Gong SP, Fang SY, Hsu YY (2007): Functional MRI of conventional and anomalous metaphors in Mandarin Chinese. *Brain Lang* 100:163–171.
- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005): Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166:559–571.
- Badre D (2008): Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn Sci* 12:193–200.
- Badre D, D’Esposito M (2009): Is the rostro-caudal axis of the frontal lobe hierarchical? *Nat Rev Neurosci* 10:659–669.
- Badre D, Poldrack RA, Pare-Blagoev EJ, Insler RZ, Wagner AD (2005): Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47:907–918.
- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004a): Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nat Neurosci* 7:1190–1192.
- Beauchamp MS, Lee KE, Argall BD, Martin A (2004b): Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41:809–823.
- Bookheimer S (2002): Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Ann Rev Neurosci* 25:151–188.
- Bunge SA, Wendelken C, Badre D, Wagner AD (2005): Analogical reasoning and prefrontal cortex: Evidence for separable retrieval and integration mechanisms. *Cereb Cortex* 15:239–249.
- Bunge SA, Helskog EH, Wendelken C (2009): Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *Neuroimage* 46:338–342.
- Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E (2004): Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *J Cogn Neurosci* 16:805–816.
- Calvert GA (2001): Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123.
- Calvert GA, Thesen T (2004): Multisensory integration: Methodological approaches and emerging principles in the human brain. *J Physiol Paris* 98:191–205.
- Dick AS, Goldin-Meadow S, Hasson U, Skipper JI, Small SL (2009): Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum Brain Mapp* 30:3509–3526.
- 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* 25:1325–1335.
- Eviatar Z, Just MA (2006): Brain correlates of discourse processing: An fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia* 44:2348–2359.
- Feyereisen P, Van de Wiele M, Dubois F (1988): The meaning of gestures: What can be understood without speech? *Curr Psychol Cogn* 8:3–25.
- Friederici AD, Ruschemeyer SA, 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* 7:30–40.
- Gold BT, Balota DA, Kirchoff BA, Buckner RL (2005): Common and dissociable activation patterns associated with controlled semantic and phonological processing: Evidence from FMRI adaptation. *Cereb Cortex* 15:1438–1450.
- Green AE, Fugelsang JA, Kraemer DJ, Shamosh NA, Dunbar KN (2006): Frontopolar cortex mediates abstract integration in analogy. *Brain Res* 1096:125–137.
- Green A, Straube B, Weis S, Jansen A, Willmes K, Konrad K, Kircher T (2009): Neural integration of iconic and unrelated coverbal gestures: A functional MRI study. *Human Brain Mapp* 30:3309–3324.
- Hagoort P, Baggio G, Willems RM (2009): Semantic unification. In: Gazzaniga MS, editor. *The Cognitive Neurosciences IV*. Cambridge: MIT Press pp 819–836.
- Hein G, Knight RT (2008). Superior temporal sulcus—It’s my area: Or is it? *J Cogn Neurosci* 20:2125–2136.

- Hein G, Doehrmann O, Muller NG, Kaiser J, Muckli L, Naumer MJ (2007): Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *J Neurosci* 27:7881–7887.
- Hocking J, Price CJ (2008): The role of the posterior superior temporal sulcus in audiovisual processing. *Cereb Cortex* 18:2439–2449.
- Hoening K, Scheef L (2009): Neural correlates of semantic ambiguity processing during context verification. *Neuroimage* 45:1009–1019.
- Holle H, Gunter TC, Ruschemeyer SA, Hennenlotter A, Iacoboni M (2008): Neural correlates of the processing of co-speech gestures. *Neuroimage* 39:2010–2024.
- Hubbard AL, Wilson SM, Calla DE, Dapretto M (2009): Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapp* 30:1028–1037.
- January D, Trueswell JC, Thompson-Schill SL (2009): Co-localization of stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing. *J Cogn Neurosci* 21:2434–2444.
- Kable JW, Chatterjee A (2006): Specificity of action representations in the lateral occipitotemporal cortex. *J Cogn Neurosci* 18:1498–1517.
- Kable JW, Kan IP, Wilson A, Thompson-Schill SL, Chatterjee A (2005): Conceptual representations of action in the lateral temporal cortex. *J Cogn Neurosci* 17:1855–1870.
- Kemmotsu N, Villalobos ME, Gaffrey MS, Courchesne E, Muller RA (2005): Activity and functional connectivity of inferior frontal cortex associated with response conflict. *Brain Res Cogn Brain Res* 24:335–342.
- 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* 47:169–179.
- 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* 15:272–293.
- Kuperberg GR, Sitnikova T, Lakshmanan BM (2008): Neuroanatomical distinctions within the semantic system during sentence comprehension: Evidence from functional magnetic resonance imaging. *Neuroimage* 40:367–388.
- Lee SS, Dapretto M (2006): Metaphorical vs. literal word meanings: fMRI evidence against a selective role of the right hemisphere. *Neuroimage* 29:536–544.
- Lewis JW, Brefczynski JA, Phinney RE, Janik JJ, DeYoe EA (2005): Distinct cortical pathways for processing tool versus animal sounds. *J Neurosci* 25:5148–5158.
- Lewis JW (2006): Cortical networks related to human use of tools. *Neuroscientist* 12:211–231.
- Luo Q, Perry C, Peng D, Xu D, Ding G, Xu S (2003): The neural substrate of analogical reasoning: An fMRI study. *Brain Res Cogn Brain Res* 17:527–534.
- Macmillan NA, Creelman CD. 1991: *Detection Theory: A User's Guide*. New York: Cambridge University Press.
- McNeill D. 1992. *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.
- McNeill D. 2005. *Gesture and Thought*. Chicago: University of Chicago Press.
- Moss HE, Abdallah S, Fletcher P, Bright P, Pilgrim L, Acres K, Tyler LK (2005): Selecting among competing alternatives: Selection and retrieval in the left inferior frontal gyrus. *Cereb Cortex* 15:1723–1735.
- Naumer MJ, Doehrmann O, Muller NG, Muckli L, Kaiser J, Hein G (2008). Cortical plasticity of audio-visual object representations. *Cereb Cortex* 19:1641–1653.
- Novick JM, Trueswell JC, Thompson-Schill SL (2005): Cognitive control and parsing: Re-examining the role of Broca's area in sentence comprehension. *Cogn Affect Behav Neurosci* 5:263–281.
- Nichols T, Brett M, Andersson J, Wager T, Poline J (2005): Valid conjunction inference with the minimum statistic. *Neuroimage* 25:653–660.
- Oldfield RC (1971): The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Rapp AM, Leube DT, Erb M, Grodd W, Kircher TJJ (2004): Neural correlates of metaphor processing. *Brain Res Cogn Brain Res* 20:395–402.
- Rapp AM, Leube DT, Erb M, Grodd W, Kircher TT (2007): Laterality in metaphor processing: Lack of evidence from functional magnetic resonance imaging for the right hemisphere theory. *Brain Lang* 100:142–149.
- Sætrevik B, Specht K (2009): Cognitive conflict and inhibition in primed dichotic listening. *Brain Cogn* 71:20–25.
- Shalom DB, Poeppel D (2008): Functional anatomic models of language: Assembling the pieces. *Neuroscientist* 14:119–127.
- Slotnick SD, Moo LR, Segal JB, Hart J (2003): Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cogn Brain Res* 17:75–82.
- Slotnick SD, Schacter DL (2004): A sensory signature that distinguishes true from false memories. *Nat Neurosci* 7:664–672.
- 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* 21:821–836.
- Straube B, Green A, Jansen A, Chatterjee A, Kircher T (2010). Social cues, mentalizing and the neural processing of speech accompanied by gestures. *Neuropsychologia* 48:382–393.
- Stringaris AK, Medford NC, Giampietro V, Brammer MJ, David AS (2007): Deriving meaning: Distinct neural mechanisms for metaphoric, literal, and nonmeaningful sentences. *Brain Lang* 100:150–162.
- Thompson-Schill SL, Bedny M, Goldberg RF (2005): The frontal lobes and the regulation of mental activity. *Curr Opin Neurobiol* 15:219–224.
- Wendelken C, Nakhabenko D, Donohue SE, Carter CS, Bunge SA (2008): "Brain is to thought as stomach is to ??": Investigating the role of rostralateral prefrontal cortex in relational reasoning. *J Cogn Neurosci* 20:682–693.
- Willems RM, Hagoort P (2007): Neural evidence for the interplay between language, gesture, and action: A review. *Brain Lang* 101:278–289.
- Willems RM, Ozyurek A, Hagoort P (2007): When language meets action: The neural integration of gesture and speech. *Cereb Cortex* 17:2322–2333.
- Willems RM, Ozyürek A, Hagoort P (2009): Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage* 47:1992–2004.