Neural Integration of Iconic and Unrelated Coverbal Gestures: A Functional MRI Study

Antonia Green,^{1,2*} Benjamin Straube,^{1,2} Susanne Weis,³ Andreas Jansen,^{2,4}
Klaus Willmes,⁵ Kerstin Konrad,⁶ and Tilo Kircher⁴

¹Department of Psychiatry and Psychotherapy—Section Experimental Psychopathology,
RWTH Aachen University, Aachen, Germany

²Interdisciplinary Centre for Clinical Research “Biomat.,” RWTH Aachen University,
Aachen, Germany

³Department of Neurology—Section Clinical Neuropsychology, RWTH Aachen University,
Aachen, Germany

⁴Department of Psychiatry und Psychotherapy, Philipps University Marburg, Marburg, Germany

⁵Department of Neurology—Section Neuropsychology, RWTH Aachen University, Aachen, Germany

⁶Department of Child and Adolescent Psychiatry and Psychotherapy—Section Child Neuropsychology,
RWTH Aachen University, Aachen, Germany

Abstract: Gestures are an important part of interpersonal communication, for example by illustrating physical properties of speech contents (e.g., “the ball is round”). The meaning of these so-called iconic gestures is strongly intertwined with speech. We investigated the neural correlates of the semantic integration for verbal and gestural information. Participants watched short videos of five speech and gesture conditions performed by an actor, including variation of language (familiar German vs. unfamiliar Russian), variation of gesture (iconic vs. unrelated), as well as isolated familiar language, while brain activation was measured using functional magnetic resonance imaging. For familiar speech with either of both gesture types contrasted to Russian speech-gesture pairs, activation increases were observed at the left temporo-occipital junction. Apart from this shared location, speech with iconic gestures exclusively engaged left occipital areas, whereas speech with unrelated gestures activated bilateral parietal and posterior temporal regions. Our results demonstrate that the processing of speech with speech-related versus speech-unrelated gestures occurs in two distinct but partly overlapping networks. The distinct processing streams (visual versus linguistic/spatial) are interpreted in terms of “auxiliary systems” allowing the integration of speech and gesture in the left temporo-occipital region. *Hum Brain Mapp* 30:3309–3324, 2009. © 2009 Wiley-Liss, Inc.

Key words: fMRI; perception; language; semantic processing; multimodal integration

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: Interdisciplinary Centre for Clinical Research “BIOMAT” within the Faculty of Medicine at the RWTH Aachen University; Contract grant number: VV N68-e; Contract grant sponsor: Deutsche Forschungsgemeinschaft (DFG); Contract grant number: IRTG 1328.

*Correspondence to: Antonia Green, Department of Psychiatry und Psychotherapy, Philipps-University Marburg, Rudolf-

Bultmann-Straße 8, D-35039 Marburg, Germany.
E-mail: psychiat@med.uni-marburg.de

Received for publication 14 July 2008; Revised 31 December 2008;
Accepted 20 January 2009

DOI: 10.1002/hbm.20753

Published online 6 April 2009 in Wiley InterScience (www.interscience.wiley.com).

INTRODUCTION

Everyone uses hands and arms to transmit information using gestures. Most prevalent are iconic gestures, sharing a formal relationship with the co-occurring speech content. Iconic gestures illustrate forms, shapes, events or actions that are the topic of the simultaneously occurring speech. The phrase “The fisherman caught a huge fish,” for example, is often accompanied by a gesture indicating the size of the fish, which helps the listener to understand how incredibly big the prey was [McNeill, 1992].

How precisely this enhanced understanding is achieved is still unclear. While speech reveals information only successively, gesture can convey multifaceted information at one time [Kita and Ozyürek, 2003; McNeill, 1992]. Thus, to combine speech and gestures into one representation gestures have to be integrated into the successively unfolding interpretation of speech semantics. Since the meaning of iconic gestures is not fixed but has to be constructed online by the listener according to the sentence content, interaction of speech and gesture processing is required on a semantic-conceptual level, as previous studies using event-related potentials have demonstrated [Holle and Gunter, 2007; Ozyürek et al., 2007; Wu and Coulson, 2007]. This interaction is distinct from basic sensory level processing as it has been studied in previous multisensory studies [for a review see Calvert, 2001]. Although information transmitted by gestures is different from speech, both modalities are tightly intertwined [McNeill, 1992]: Speech and gestures usually transmit at least a similar meaning, the most meaningful part of the gesture (stroke) temporally aligned with the respective speech segment and both aiming at communicating a message. Hence, a processing mechanism integrating both sources of information seems rather likely.

Findings about the neural correlates of isolated processing of speech or gestures converge on the assumption that understanding language as well as gestures relies on partly overlapping brain networks [for a review see Willems and Hagoort, 2007]. Speech-gesture interactions so far have only been targeted by a few recent fMRI studies that used different kinds of paradigms: either mismatch manipulations [gestures that were incongruent with the utterance; Willems et al., 2007], a disambiguation paradigm [gestures clarifying the meaning of ambiguous words; Holle et al., 2008] or a memory paradigm [effects of speech-gesture relatedness on subsequent memory performance; Straube et al., in press]. Across these studies, brain activations were reported in the left inferior frontal gyrus (IFG), inferior parietal cortex, posterior temporal regions (superior temporal sulcus, middle temporal gyrus) and in the precentral gyrus. Activity in the left IFG was repeatedly found for speech-gesture pairs that were more difficult to integrate (mismatching speech-gesture pairs or abstract sentence contents), temporal activations have been related to semantic integration, and inferior parietal involvement has been related to action processing, some-

times interpreted in terms of the putative human mirror neuron system (MNS) [Rizzolatti and Craighero, 2004]. This system is meant to determine the goals of actions by an observation-execution matching process [Craighero et al., 2007] showing stronger activations when the initial goal hypothesis (internal motor simulation) of an observed action is not matched by the visual input and initiating new simulation cycles. To summarize, due to the different paradigms it cannot be clearly stated which regions in the brain are implicated in natural speech-gesture integration.

In the present study, we investigated the neural networks engaged in the integration of speech and iconic gestures. By “integration” we refer to implicitly initiated cognitive processes of combining semantic audiovisual information into one representation. We assume that this leads to increased processing demands which should be reflected in activation increases in brain regions involved in the processing of gestures with familiar speech as opposed to gestures with unfamiliar speech. This increased activation is hypothesized to reflect the creation of a semantic connection of both modalities. Subjects were presented with short video clips of an actor speaking German (comprehensible) or Russian sentences (incomprehensible to our subjects who had no knowledge of Russian) that were either presented in isolation, were accompanied by iconic gestures, or by gestures that were unrelated to the sentence. The Russian language and speech-unrelated gesture conditions were used as control conditions in order to be able to present both modalities, but preventing semantic processing in the respective irrelevant modality.

We were interested in two questions: first, we wanted to identify brain areas activated by the natural semantic integration of speech and iconic gestures. Second, we examined whether the same cerebral network is engaged in the processing of speech with speech-unrelated gestures.

We hypothesized that integration of congruent iconic speech-gesture information would be reflected in left posterior temporal activations [Holle et al., 2008; Straube et al., in press]. For unrelated gestures we expected that because of less congruency between speech and gestures mainly parietal regions would be involved, most likely reflecting the mapping of irrelevant complex movements onto speech [Willems et al., 2007].

MATERIALS AND METHODS

Participants

Sixteen healthy male subjects were included in the study (mean age 28.8 ± 8.3 years, range 23–55 years). All participants were right handed [Oldfield, 1971], native German speakers and had no knowledge of Russian. The subjects had normal or corrected-to-normal vision, none reported hearing deficits. Exclusion criteria were a history of relevant medical or psychiatric illness of the participant himself or in his first-degree relatives. All subjects gave

Figure 1.

Design with examples of the video stimuli. The stimulus material consists of video clips of an actor speaking and performing gestures (exemplary screenshots). Speech bubbles (translations of the original German sentence “Der Angler hat einen großen Fisch gefangen”) and arrows (indicating the movement of the

hands) are inserted for illustrative purposes only. Note the dark- and light-colored spots on the actor’s sweater that were used for the control task. The original stimuli were in color (see examples in the Supporting Information).

written informed consent before participation in the study. The study was approved by the local ethics committee.

Stimulus Material

The stimulus production is described in more detail in Straube et al. [in press] for gestures in the context of abstract speech contents (metaphoric coverbal gestures), here only a shorter description is given. In the current study though, different stimuli with concrete speech contents were used, but these were produced accordingly.

The stimulus material consisted of short video clips showing an actor who performed combinations of speech (German and Russian) and gestures (iconic and unrelated gestures) (see Fig. 1). Initially, a set of 1,296 (162×8) videos with eight conditions was created: (1) German sentences with corresponding iconic gestures [GI], (2) German sentences with unrelated gestures [GU], (3) Russian sentences with corresponding iconic gestures [RI], (4) Russian sentences with unrelated gestures [RU], (5) German sentences without gestures [G], (6) Russian sentences without gestures, (7) meaningful iconic gestures without speech [I], and (8) unrelated gestures without speech [U].

This study focuses on five stimulus types (see Fig. 1 and examples in the Supporting Information): (1) German sentences with corresponding iconic gestures [GI], (2) German sentences with unrelated gestures [GU], (3) Russian sentences with corresponding iconic gestures [RI], (4) Russian sentences with unrelated gestures [RU], and (5) German sentences without gestures [G].

We constructed German sentences that contained only one element which could be illustrated by a gesture. The gesture had to match McNeill’s “iconic gestures” definition in that it illustrates the form, size or movement of something concrete that is mentioned in speech [McNeill, 1992]. The sentences had the same length of five to eight words and a similar grammatical form (subject–predicate–object). All sentences were translated into Russian, order to present natural speech without understandable semantics. In addition, we developed “unrelated” gestures as a control condition, in order to produce speech–gesture pairs without clear-cut gesture–speech mismatches but that contain movements that are as complex (e.g., gesture direction and extent), smooth and vivid as the iconic gestures. The only difference was that unrelated gestures had obvious meaning when presented in isolation and had only a very weak relation to the sentence context (for details see Supporting Information Material). A rating procedure (see

below and in the Supporting Information) has proven that in fact our unrelated gestures did not contain any clear-cut semantic information and that they differed significantly in semantic strength from the iconic gestures.

A male bilingual actor was instructed to produce each utterance together with an iconic or unrelated gesture in a way that he felt as the originator of the gestures. Thus, the synchrony of speech and gesture was determined by the actor and was left unchanged during the editing process. It is important to note that the unrelated gestures were only roughly choreographed, but not previously scripted. They rather were developed in collaboration with the actor, often derived on the basis of already used iconic gestures and trained until they looked and felt like intrinsically produced spontaneous gestures. By doing so we ensured that the gestures evolved at the correct moment and were as smooth and dynamic as the iconic gestures. Only when a specific GI sentence was recorded together with an iconic gesture in a natural way the respective control conditions were produced successively. This was done to keep all item characteristics constant (e.g., sentence duration or movement complexity), regardless of the manipulated factors of language (German, Russian, no speech) and gesture (iconic, unrelated, no gesture). Before and after the utterance the actor stood with hands hanging comfortably. Each clip had a duration of 5,000 ms, including at least 500 ms before and after the scene, where the actor neither spoke nor moved. This was done to account for variable utterance lengths to get standardized clips.

The recorded sentences presented here had an average speech duration of 2,330 ms ($SD = 390$ ms) and an average gesture duration of 2,700 ms ($SD = 450$ ms) and did not differ between conditions (see Table S-II, Supporting Information Material).

Based upon ratings of understandability, imageability and naturalness (see Supporting Information Material, Table S-I) as well as upon parameters such as movement characteristics, pantomimic content, transitivity or handedness, we chose a set of 1,024 video clips (128 German sentences with iconic gestures and their counterparts in the other seven conditions) from the initial set of 1,296 clips as stimuli for the fMRI experiment. Stimuli were divided into four sets in order to present each participant with 256 clips during the scanning procedure (32 items per condition), counterbalanced across subjects. Across subjects, each item was presented in all eight conditions but a single participant only saw complementary derivatives of one item, i.e. the same sentence or gesture information was only seen once per participant. This was done to prevent speech or gesture repetition effects. Again, all parameters listed above were used for an equal assignment of video clips to the four experimental sets, to avoid set-related between-subject differences.

Previous research has shown that coverbal gestures are not only semantically but also temporally aligned with the corresponding utterance. We assured this natural syn-

chrony by making the actor initiate the gestures and additionally checked for it when postprocessing the video clips. Each sentence contained only one element that could be illustrated, which was intuitively done by the actor. During postprocessing we checked in the GI condition with which word the gesture stroke (peak movement) coincided and then chose the end of this word as the time point of highest semantic correspondence between speech and the gesture stroke, assuming that the stroke of the gesture precedes or ends at, but does not follow the phonological peak syllable of speech [McNeill, 1992]. For example, for the sentence "The house has a vaulted roof" the temporal alignment was marked at the end of the word ("vaulted") coinciding with and corresponding to the iconic gesture (form representation) in the GI condition. The time of alignment was then transferred to the Russian equivalent (condition RI). This was possible because of the standardized timing, form and structure of each of the different conditions corresponding to an item. Accordingly, the time of alignment was determined for the GU and G conditions (end of "vaulted") and transferred to the Russian counterparts. The concordance between GI and GU conditions was checked and revealed no differences, the times of alignment occurred on average 1.17 s (GI, $SD = 0.51$ second)/1.19 seconds (GU, $SD = 0.50$ second) after the actual gesture onset and 1.53 s (GI, $SD = 0.53$ second)/1.54 seconds (GU, $SD = 0.58$ second) before the actual gesture offset. The time of alignment representing the stroke-speech synchrony occurred on average 2,150 ms ($SD = 535$ ms) after the video start (1,650 ms after speech onset) (see Table S-II, Supporting Information Material) and was used for the modulation of events in the event-related fMRI analysis.

Experimental Design and Procedure

An experimental session comprised 256 trials (32 for each condition) and consisted of four 11-minute blocks. Each block contained 64 trials with a matched number of items from each condition. The stimuli were presented in an event-related design in pseudo-randomized order and counterbalanced across subjects. Each clip was followed by a fixation cross on grey background with a variable duration of 3,750–6,750 ms (average: 5,000 ms).

During scanning participants were instructed to watch the videos and to indicate via left hand key presses at the beginning of each video whether the spot displayed on the actor's sweater was light or dark colored. This task was chosen to focus participants' attention on the middle of the screen and enabled us to investigate implicit speech and gesture processing without possible instruction-related attention biases. Performance rates and reaction times were recorded. Before scanning, each participant received at least 10 practice trials outside the scanner, which were different from those used in the main experiment. Before

the experiment started, the volume of the videos was individually adjusted so that the clips were clearly audible.

Fifteen minutes after scanning an unannounced recognition paradigm was used to control for participants' attention and to examine the influence of gestures on memory performance. All videos of the German-iconic and German-unrelated conditions (32 each) and half of the isolated speech condition (16) were presented in random order, together with an equal number of new items for each of the three conditions (altogether 160 videos). Participants had to indicate via key press whether they had seen that clip before or not. Memory data of three participants are missing for technical reasons.

MRI Data Acquisition

The video clips were presented via MR-compatible video goggles (stereoscopic display with up to $1,024 \times 768$ pixel resolution) and nonmagnetic headphones. Furthermore, participants wore ear plugs, which act as an additional low-pass filter.

All MRI data were acquired on a Philips Achieva 3T scanner. Functional images were acquired using a T2*-weighted echo planar image sequence (TR = 2 seconds, TE = 30 ms, flip angle 90° , slice thickness 3.5 mm with a 0.3-mm interslice gap, 64×64 matrix, FoV 240 mm, in-plane resolution 3.5×3.5 mm, 31 axial slices orientated parallel to the AC-PC line covering the whole brain). Four runs of 330 volumes were acquired during the experiment. The onset of each trial was synchronized to a scanner pulse. Additionally, an anatomical scan was acquired for each participant using a high resolution T1-weighted 3D-sequence consisting of 180 sagittal slices (TR = 9,863 ms, TE = 4.59 ms, FoV = 256 mm, slice thickness 1 mm, interslice gap = 1 mm).

MRI Data Analysis

SPM2 (www.fil.ion.ucl.ac.uk) standard routines and templates were used for analysis of fMRI data. After discarding the first five volumes to minimize T1-saturation effects, all images were spatially and temporally realigned, normalized (resulting voxel size $4 \times 4 \times 4$ mm³), smoothed (8 mm isotropic Gaussian filter) and high-pass filtered (cut-off period 128 seconds).

Statistical analysis was performed in a two-level, mixed-effects procedure. At the first level, single-subject BOLD responses were modeled by a design matrix comprising the onsets of each event (i.e., time of alignment, see stimulus construction) of all eight experimental conditions. The hemodynamic response was modeled by the canonical hemodynamic response function (HRF) and its temporal derivative. The volume of interest was restricted to grey matter voxels by use of an inclusive mask created from the segmentation of the standard brain template. Parameter estimate (β -) images for the HRF were calculated for

each condition and each subject. Direct contrasts between events (GI > RI, GI > G, GI > I, GU > RU, GU > G, GU > U, GU > GI) were computed per participant. At the second level, a random-effects group analysis was performed by entering corresponding contrast images of the first level for each participant into one-sample *t*-tests to compute statistical parametric maps for the above contrasts. All difference contrasts were inclusively masked by their minuends to ensure that only differences with respect to the activations of the first condition are evaluated. Further analyses and the contrasts of interest are described below in more detail.

We chose to employ Monte-Carlo simulation of the brain volume to establish an appropriate voxel contiguity threshold [Slotnick and Schacter, 2004]. This correction has the advantage of higher sensitivity 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. To implement such an approach, we ran a Monte-Carlo simulation to model the brain volume (<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, FWHM of 6.5 mm—this value was estimated using the *t*-statistic map associated with the contrast of interest [GI > RI $\cap$ GI > G $\cap$ GU > RU $\cap$ GU > G]; similar procedures have been used previously to estimate fMRI spatial correlation, e.g., see Katanaoda et al., 2002; Ross and Slotnick, 2008; Zarahn et al., 1997]. An individual voxel threshold was then applied to achieve the assumed voxel-wise Type I error rate ($P < 0.05$). The probability of observing a given cluster extent was computed across iterations under $P < 0.05$ (corrected for multiple comparisons). In the present study, this translated to a minimum cluster extent threshold of 23 contiguous resampled voxels. In order to basically demonstrate the expected regions we present all contrasts at the same threshold. The reported voxel coordinates of activation peaks are located in MNI space. For the anatomical localization the functional data were referenced to probabilistic cytoarchitectonic maps [Eickhoff et al., 2005].

Contrasts of Interest

A common way of testing for semantic processing is the manipulation of semantic fit, here between a sentence and a gesture [e.g., Willems et al., 2007]. Incorrect or mismatching information is thought to increase semantic integration load, revealing areas that are involved in integration processes if contrasted against matching pairs [e.g., Friederici et al., 2003; Kuperberg et al., 2000; Kutas and Hillyard, 1980]. To compare the results of the present study with

previous results, we first calculated the difference contrast between unrelated and iconic gestures in combination with German speech [$GU > GI$].

We assumed that the temporospatial co-occurrence of speech and gestures results in some sort of integration processes not only for iconic but also for unrelated gestures. Unrelated or “mismatching” gestures, however, may result in unnatural processing leading not only to stronger activations in involved brain areas but also to activations in regions that are not engaged in natural speech-gesture processing. In addition, if both related and unrelated gestures are integrated with speech in a similar way, brain areas common to both processes will not be detected by such mismatch contrasts. Finally, it cannot be ruled out that activations revealed by such an analysis have resulted from or at least have been influenced by error detection or mismatch processing. Instead, we were interested in the neural correlates of naturally and implicitly occurring speech-gesture integration processes which we revealed by a stepwise analysis that gradually isolated the process we were interested in. In the context of this article we define integration as follows: semantic integration can occur only in combination with understandable language (here German, G), perhaps not only for iconic (I) but also for unrelated gestures (U). With unfamiliar language (like Russian for our subjects, R) it is impossible to create a common representation although both modalities are presented. Thus, integration processes in the familiar language conditions should be reflected in additional activations as compared with the unfamiliar language conditions. Therefore, we subtracted the Russian gesture conditions from the respective German gesture conditions (contrast 1: [$GI > RI$], [$GU > RU$]), revealing not only the neural correlates of integration processes but also those of more fundamental processing of speech semantics. In order to further subtract the activations related to semantic processing (in which we were not interested), we incorporated the isolated familiar language condition (contrast 2: [$GI > G$], [$GU > G$]) in the analyses. In a next step contrasts 1 and 2 were entered into conjunction analyses that should only reveal activations related to integration processes. This procedure resulted in an iconic conjunction analysis [$GI > RI \cap GI > G$] and an unrelated conjunction analysis [$GU > RU \cap GU > G$], both testing for independent significant effects compared at the same threshold [using the minimum statistic compared with the conjunction null, see Nichols et al., 2005]. Finally, we wanted to know whether there are similarities or differences between the processing of iconic and unrelated speech-gesture pairs which might help in clarifying the specific functions of activated regions. Specifically, this question targeted the previously isolated processes that are most likely related to integration sparing motion or auditory processing. To address the question of overlap between areas involved in integration processes for iconic as well as for unrelated gestures we entered both conjunctions into a first-order conjunction analysis [$GI > RI \cap GI > G \cap GU > RU \cap GU > G$]. Logi-

cally, this procedure equals a second-order conjunction analysis of the two previous conjunction analyses [$(GI > RI \cap GI > G) \cap (GU > RU \cap GU > G)$]. Exclusive masking (mask threshold $P < 0.05$ uncorrected) was used to identify voxels where effects were not shared between the two conjunctions, showing the distinctness of iconic versus unrelated speech-gesture interaction sites.

To enable interpretation of regions detected by the overlap analysis, a more classical analysis from the domain of multisensory integration was added. It is designed to show enhanced responses to the bimodal audiovisual stimuli [GI , GU] relative to either auditory [G] or visual [I , U] stimuli alone ($[0 < G < GI > I > 0]$ and $[0 < G < GU > U > 0]$); [see Beauchamp, 2005; Hein et al., 2007]. Small-volume correction of these results was computed on a 10-mm sphere around the coordinates localized by the overlap conjunction analysis (small-volume correction at $P < 0.05$).

RESULTS

Behavioral Results

The average reaction time for the control task (“indicate the color of the spot on the actor’s sweater”) did not differ across colors and conditions (color: $F_{1,15} = 0.506$, $P = 0.488$; condition: $F_{4,60} = 0.604$, $P = 0.604$; interaction: $F_{4,60} = 1.256$, $P = 0.301$; within-subjects two-factorial ANOVA; mean = 1.23 seconds, SD = 0.94). Participants showed an average accuracy rate of 99% which did not differ across conditions ($F_{4,60} = 0.273$, $P = 0.841$, within-subjects ANOVA). Thus, the attention control task indicates that participants did pay attention to the video clips.

In contrast to the performance in the control task, the subsequent recognition performance of our participants was significantly influenced by the item condition ($F_{2,24} = 12.336$, $P < 0.005$). This was caused by better performance for the bimodal familiar speech plus gesture conditions (GI : 50.9% correct; GU : 55.5% correct) as compared with the isolated familiar speech condition (G : 34.3% correct; both $P < 0.05$). As a tendency GU items were better recalled than GI items, but this difference was not significant ($P < 0.129$). Taken together these results suggest that both bimodal conditions led to better encoding, presumably through offering an opportunity for integration.

fMRI Results

The fMRI results are presented as described in the Methods section, first showing the results of the direct contrast between unrelated and iconic gestures. Second, the conjunction analyses that were designed to reveal activations related to natural semantic integration processes are gradually developed and finally checked for overlaps and differences of iconic and unrelated speech-gesture processing.

TABLE I. Brain activations for unrelated versus iconic gestures in combination with familiar language

	Region	Cluster extension	<i>x</i>	<i>y</i>	<i>z</i>	Extent	<i>t</i> -Value
<i>German unrelated > German iconic [GU > GI]</i>							
L	Inferior parietal lobule	Precuneus, SPL, IPC, hIP, post- and precentral gyrus	−44	−36	40	217	4.37
L	Inferior frontal gyrus	Precentral gyrus, MFG	−60	16	24	195	3.84
R	Supramarginal gyrus	IPC, hIP, angular gyrus, STG	52	−36	36	142	3.64
R	Inferior frontal gyrus		48	32	12	136	3.68
R/L	Supplementary motor area	SFG, SMG	8	20	56	64	2.88
R	Paracentral lobule	Postcentral gyrus, SPL, precuneus	8	−36	56	55	3.32
L	Middle temporal gyrus	ITG, IOG	−60	−60	−4	45	3.67
R	Middle frontal gyrus		44	12	48	29	2.59
L	Hippocampus	Putamen	−20	−12	−16	23	2.82

Stereotaxic coordinates in MNI space and *t*-values of the foci of maximum activation ($P < 0.05$ corrected). hIP, human intraparietal area; IOG, inferior occipital gyrus; IPC, inferior parietal cortex; ITG, inferior temporal gyrus; MFG, middle frontal gyrus; SFG, superior frontal gyrus; SMG, superior medial gyrus; SPL, superior parietal lobule; STG, superior temporal gyrus.

Unrelated Versus Iconic Gestures in the Context of Familiar Language

The direct comparison of unrelated versus iconic speech-gesture pairs [GU > GI] resulted in large activation clusters in the IFG (BA 44, 45) and supplementary motor areas (SMA, BA 6) bilaterally, in the left inferior parietal lobule, inferior temporal cortex and hippocampus as well as in the right supramarginal gyrus, postcentral gyrus and middle frontal gyrus (Table I and Fig. 2).

Neural Correlates of Speech-Gesture Integration Processes for Iconic and Unrelated Gestures

We start building the conjunction analyses by revealing the neural correlates of integration processes confounded by more basic semantic processing for iconic and unrelated gestures, respectively (contrast 1). For iconic gestures [GI > RI] we found the largest cluster of activation along the left middle temporal gyrus (MTG) spreading into the IFG. Other areas of activation included the right MTG, left inferior parietal lobule, left superior frontal and medial gyrus and left hippocampus. For unrelated gestures [GU > RU] we observed similar activations; however, with additional activity in left middle frontal and right inferior parietal regions (Table II and Fig. 3A).

In the next step, a contrast that controls for activations related to general language processing is computed for both kinds of gestures (contrast 2). For iconic gestures [GI > G] the largest clusters of activation were found in left and right occipito-temporal regions. Other areas of activation included the right superior temporal gyrus and left hippocampus. For unrelated gestures [GU > G] similar activation patterns were observed with bilateral parieto-temporal regions being additionally involved (Table III and Fig. 3B).

To proceed, the conjunctions of contrast 1 and 2 were computed for iconic [GI > RI $\cap$ GI > G] and unrelated gestures [GU > RU $\cap$ GU > G], respectively. These conjunctions were designed to reveal the neural correlates of semantic integration uninfluenced by pure speech processing. For both conjunctions the left MTG was significantly activated in combination with the familiar language. For iconic gestures this cluster extended more occipitally, for the unrelated gestures, more temporally. For the unrelated gesture condition additional activations were found in the left MTG, inferior occipital gyrus, inferior parietal lobule and postcentral gyrus (BA 2) and in the right inferior temporal gyrus and inferior parietal lobule (Table IV and Fig. 4A, with iconic gestures in green, unrelated gestures in red).

The overlap of the areas involved in integration processes for iconic and unrelated gestures was statistically confirmed by calculating a conjunction of all previous contrasts [GI > RI $\cap$ GI > G $\cap$ GU > RU $\cap$ GU > G] and revealed a cluster in the left posterior MTG at the TO junction (see Table IV and Fig. 4A, with overlap in yellow).

Figure 2.

Unrelated versus related gestures. Brain areas stronger activated for familiar language with unrelated gestures compared with familiar language with iconic gestures [GU > GI]. For coordinates and statistics see Table I. Map is thresholded at $P < 0.05$ (corrected) (GI = German iconic, GU = German unrelated).

TABLE II. Brain activations for gestures with familiar language versus gestures with unfamiliar language

	Region	Cluster extension	<i>x</i>	<i>y</i>	<i>z</i>	Extent	<i>t</i> -Value
<i>German iconic > Russian iconic [GI > RI]</i>							
L	Middle occipital gyrus	IFG, MTG, fusiform gyrus, ITG, HC	-40	-72	8	528	5.39
R	Medial temporal pole		44	12	-28	36	3.85
R	Middle temporal gyrus	STG	60	-44	-8	34	3.43
L	Postcentral gyrus	IPL, supramarginal gyrus	-60	-20	36	27	2.40
R	Parahippocampal gyrus		12	-8	-16	25	3.10
<i>German unrelated > Russian unrelated [GU > RU]</i>							
L	Middle temporal gyrus	IOG	-56	-8	-20	318	6.61
L	Inferior frontal gyrus		-44	28	0	259	5.38
R	Inferior temporal gyrus	MTG	56	-52	-8	85	3.76
L	Inferior parietal lobule	Post- and precentral gyrus	-32	-44	44	62	3.39
L	Precentral gyrus	Postcentral gyrus, MFG	-44	4	48	62	2.99
R	Middle temporal gyrus		56	0	-24	48	5.43
R	Inferior parietal lobule		36	-52	48	36	2.90
R	Postcentral gyrus	Supramarginal gyrus	60	-4	36	29	2.63

Stereotaxic coordinates in MNI space and *t*-values of the foci of maximum activation ($P < 0.05$ corrected). HC, hippocampus; IFG, inferior frontal gyrus; IOG, inferior occipital gyrus; IPL, inferior parietal lobule; ITG, inferior temporal gyrus; MFG, middle frontal gyrus; MTG, middle temporal gyrus; STG, superior temporal gyrus.

Finally, both conjunctions were masked reciprocally with each other to reveal differences of iconic and unrelated speech-gesture processing ($[GI > RI \cap GI > G]$ exclusively masked by $[GU > RU \cap GU > G]$; $[GU > RU \cap GU > G]$ exclusively masked by $[GI > RI \cap GI > G]$). Apart from the overlap calculated above all other activations observed for the iconic as well as for the unrelated conjunction analyses were shown to be unaffected by the respective other gesture condition (see Table IV).

According to probabilistic cytoarchitectonic maps there is only a small possibility (below 30%, see Table IV and Fig. 4B) that the activations revealed by the conjunction

analyses are located in area hOC5 (V5/MT+) [Eickhoff et al., 2005; Malikovic et al., 2007]. Nevertheless we cannot exclude that parts of the activation are situated within that region.

To further elucidate the function of the overlap region we conducted a multimodal integration analysis for related $[0 < G < GI > I > 0]$ and unrelated gestures $[0 < G < GU > U > 0]$, respectively, revealing areas that are strongly activated by speech-gesture stimuli than by isolated speech or gestures. According to these analyses both kinds of gestures elicited multimodal integration in the TO region (iconic speech gesture pairs: $t = 2.12$; unrelated

Figure 3.

Contrasts entered into the conjunction analyses. **(A)** Difference contrasts of gestures with familiar language (German) versus pairs with unfamiliar language (Russian), revealing semantic and integration processes, and **(B)** contrasts of gestures with familiar language (German) versus isolated German, controlling for lan-

guage processing; for iconic gestures (top) and unrelated gestures (bottom). For coordinates and statistics see Tables II and III. Maps are thresholded at $P < 0.05$ (corrected) (GI = German iconic, RI = Russian iconic, GU = German unrelated, RU = Russian unrelated, G = German without gestures).

TABLE III. Brain activations for gestures with familiar language versus isolated familiar language

Region	Cluster extension	<i>x</i>	<i>y</i>	<i>z</i>	Extent	<i>t</i> -Value
<i>German iconic > German [GI > G]</i>						
L Inferior occipital gyrus	SOG, MOG, cerebellum, fusiform gyrus, IOG	-48	-76	-4	299	6.24
R Inferior temporal gyrus	MOG, SOG, IOG, fusiform gyrus, cuneus	52	-72	-4	274	7.06
R Superior temporal gyrus		68	-24	8	33	3.52
L Hippocampus		-20	-28	-8	25	3.54
<i>German unrelated > German [GU > G]</i>						
L Inferior occipital gyrus	Supramarginal gyrus, MOG, SOG, postcentral gyrus, IPL, hIP, calcarine gyrus, fusiform gyrus, STG, MTG	-48	-76	-4	527	6.71
R Inferior temporal gyrus	IOG, MOG, ITG, STG, MTG, SOG, cuneus, lingual gyrus	44	-68	-4	292	6.31
R Angular gyrus	SPL, postcentral gyrus, IPL, hIP	32	-56	44	82	3.30

Stereotaxic coordinates in MNI space and *t*-values of the foci of maximum activation ($P < 0.05$ corrected). hIP, human intraparietal area; IOG, inferior occipital gyrus; IPL, inferior parietal lobule; ITG, inferior temporal gyrus; MOG, middle occipital gyrus; MTG, middle temporal gyrus; SOG, superior occipital gyrus; SPL, superior parietal lobule; STG, superior temporal gyrus.

speech gesture pairs: $t = 2.53$, both $P < 0.05$ corrected after small volume correction with a 10-mm sphere at $-52 -68 0$).

DISCUSSION

The goal of the present study was to reveal the neural correlates of semantic interaction processes of iconic speech-gesture pairs and to examine whether there are similar activations when gestures are unrelated to speech. Subjects saw video clips of an actor speaking sentences in a familiar (German) or unfamiliar (Russian) language that were either presented in isolation or were accompanied by iconic gestures or by gestures that were unrelated to the sentence.

As previous research on speech and gesture processing using ERPs [Ozyürek et al., 2007] and fMRI [Willems et al., 2007] has focused on incongruent speech-gesture information, we first contrasted directly the processing of unrelated gestures with iconic gestures. In line with a previous study, we found inferior frontal as well as parietal activations [Willems et al., 2007]. Willems and colleagues interpreted this pattern as reflecting stronger integration load, but we cannot rule out that these activations are a result of the processing of unnatural stimuli and rather relate to error detection processes.

The main focus of the present study, however, was the analysis of neural networks engaged in the natural integration of speech and iconic gestures. We showed that semantically related speech and iconic gestures with familiar as opposed to unfamiliar speech activates

TABLE IV. Conjunction analyses—generally involved, overlapping, and distinct regions

Region	Cluster extension	<i>x</i>	<i>y</i>	<i>z</i>	Extent	<i>t</i> -Value
<i>Conjunction analysis for iconic gestures [GI > RI ∩ GI > G]</i>						
L Middle occipital gyrus	Probability for V5/MT+: 30%	-48	-68	4	72	3.98
<i>Conjunction analysis for unrelated gestures [GU > RU ∩ GU > G]</i>						
L Middle temporal gyrus	IOG	-56	-68	0	99	3.22
R Inferior temporal gyrus		56	-64	-4	47	3.14
L Inferior parietal lobule	Postcentral gyrus	-32	-44	44	39	3.05
R Inferior parietal lobule		32	-52	44	25	2.54
<i>Overlap: Conjunction of iconic and unrelated conjunction analyses [GI > RI ∩ GI > G] ∩ [GU > RU ∩ GU > G]</i>						
L Middle temporal gyrus	Probability for V5/MT+: 20%	-52	-68	0	24	3.01
<i>Distinct: Iconic conjunction exclusively masked by unrelated conjunction [GI > RI ∩ GI > G] excl. masked by [GU > RU ∩ GU > G]</i>						
L Middle occipital gyrus	Probability for V5/MT+: 20%	-40	-76	8	48	3.94
<i>Distinct: Unrelated conjunction exclusively masked by iconic conjunction [GU > RU ∩ GU > G] excl. masked by [GI > RI ∩ GI > G]</i>						
L Middle temporal gyrus	IOG (probability for V5/MT+: 20%)	-56	-56	0	71	2.85
R Inferior temporal gyrus	Probability for V5/MT+: 20%	56	-64	-4	45	3.14
L Inferior parietal lobule	Postcentral gyrus	-32	-44	44	39	3.05
R Inferior parietal lobule		32	-52	44	25	2.54

Stereotaxic coordinates in MNI space and *t*-values of the foci of maximum activation ($P < 0.05$ corrected). IOG, inferior occipital gyrus.

Figure 4.

Results of the conjunction analyses. **(A)** Brain areas related to the integration of speech and gestures, based on the following conjunctions: $[GI > RI \cap GI > G]$ (green, iconic gestures) and $[GU > RU \cap GU > G]$ (red, unrelated gestures)—note the overlap in the left temporo-occipital junction (yellow). Parameter estimates (arbitrary units) are shown for the local maxima as

listed in Table IV. **(B)** Transverse sections through the TO region revealing activation related to both conjunction analyses (overlap). For coordinates and statistics see Table IV. Maps are thresholded at $P < 0.05$ (corrected) (GI = German iconic, RI = Russian iconic, GU = German unrelated, RU = Russian unrelated, G = German without gestures).

the left temporo-occipital (TO) junction. For unrelated gestures with familiar as opposed to unfamiliar speech similar activation increases were observed. However,

areas involved in the processing of coverbal iconic and unrelated gestures, respectively, are only partly overlapping (posterior MTG) and largely involve distinct

regions, specifically activated for either iconic (middle occipital gyrus) or unrelated gestures (bilateral temporal and parietal regions). Based on our findings we propose that both related and unrelated gestures induce an integration process that is reflected in two different processing streams converging in the left posterior temporal gyrus.

Processing of Iconic Gestures With Familiar Speech

Speech and gesture have to be integrated online into a semantic representation. To avoid problems associated with the comparison of unrelated and related gestures we investigated the processing of natural speech-gesture pairs and manipulated the understandability of the speech component. Our first hypothesis targeted the neural correlates of semantic integration processes, independent of mismatch manipulations.

The neural correlate of the integration of iconic gestures with familiar German speech was located in the left temporo-occipital (TO) junction, extending from the MTG into the superior occipital gyrus. Several studies have demonstrated the involvement of the left posterior MTG in language processing [for reviews see Démonet et al., 2005; Vigneau et al., 2006], in contextual sentence integration [Baumgärtner et al., 2002], in action representations and concepts [e.g., Kable et al., 2005; Martin and Chao, 2001] and in multisensory integration for the identification of tool or animal sounds [Beauchamp et al., 2004b; Lewis, 2006]. It has been argued that multimodal responses in the posterior MTG may reflect the formation of associations between auditory and visual features that represent the same object [Beauchamp et al., 2004b].

Similarly, activations in the left TO junction have also been demonstrated for a variety of tasks, including meaning-based paradigms as well as diverse paradigms related to visual aspects such as motion. Hickok and Poeppel [2000] argue in their review that language tasks accessing the mental lexicon (i.e., accessing meaning-based representations) rely on auditory-to-meaning interface systems in the vicinity of the left TO junction and call this region an auditory-conceptual interface, which may be part of a more widely distributed and possibly modality-specific conceptual network [Barsalou et al., 2003; Hickok and Poeppel, 2007; Tranel et al., 2008]. More generally, the left TO junction is a multimodal association area involved in semantic processing, including the mapping of visual input onto linguistic representations [Ebisch et al., 2007]. Furthermore, patients with lesions in this region show ideational apraxia, an impaired knowledge of action concepts and inadequate object use [De Renzi and Lucchelli, 1988]. Tranel et al. [1997] found in a large sample of patients with brain lesions that defective conceptual knowledge was related to lesions in the left TO junction. As 50% of the items used in this study were related to

objects and another 17% were reenacted actions without objects [which is similar to proportions observed in spontaneous speech, cf. Hadar and Krauss, 1999] it is likely that the activation of the TO junction resulted from these motion-related images.

Apart from involvement in semantic-conceptual processes, the TO junction is engaged in the processing of visual input. It was involved in the processing of motion even if static stimuli only implied motion [Kourtzi and Kanwisher, 2000] or were somehow related to visual topographical or spatial information [e.g., Mummery et al., 1998]. Thus, it seems likely that a region that is sensitive to semantic concepts as well as to visuospatial aspects integrates information from speech and gestures into one representation.

At first glance the activation of temporo-occipital regions for the integration of familiar speech and iconic gestures seems to be inconsistent with previous fMRI studies on the integration of speech and iconic gestures, mentioning instead the left IFG and superior temporal sulcus [STS] as the key regions for speech-gesture integration. These discrepancies, however, may be explained by differences in experimental paradigms. Willems et al. [2007] found activation of the left STS specifically for the language mismatch (sentences with inappropriate verb and a gesture fitting the expected verb) but not for the gesture mismatch (correct sentences with an incongruent gesture) or the double mismatch condition (sentences with inappropriate verb and a gesture fitting the inappropriate verb). This might reflect a violation of expectancies derived from speech semantics, as has been shown by Ni et al. [2000] and Kuperberg et al. [2003]. The semantic mismatch may be much more dominant in the language mismatch condition (absence of useful language information) than in the gesture mismatch condition. The apparent mismatch in speech may lead to a stronger focus on gesture semantics. These might be used to “correct” the faulty utterance and therefore they need to be integrated even more into the preceding sentence context. In the case of gesture mismatches this interaction between context and retrieval of lexical-semantic information may be less important because the speech semantics likely dominate the processing of the whole sequence. The activation of the STS in the study by Holle et al. [2008] might rather reflect the interaction between the meaning of a gesture and an ambiguous sentence. We, in contrast, used only unambiguous, natural sentences without mismatches for this analysis. Importantly, a closer look at the coordinates of the STS-clusters reported in the study by Holle et al. [2008] reveals that these activations are located in the temporo-occipital cortex, showing their strongest activation for dominant meanings in the middle occipital gyrus and for subordinate meanings in the occipitotemporal junction. Thus, our results are largely in congruence with the study of Holle et al. Additionally, studies on metaphoric gestures performed by our group also revealed

integration-related activations in the left posterior MTG [Kircher et al., 2009; Straube et al., in press].

The focus on the pSTS in integration processes mainly stems from nonhuman primate studies demonstrating converging afferents from different senses in the primate homolog of human pSTS [e.g., Seltzer and Pandya, 1978]. Further support comes from several studies associating this region with the crossmodal binding of audiovisual input [e.g., Calvert et al., 2000]. But there also are studies that did not report enhanced activation for bimodal versus unimodal stimuli [e.g., Taylor et al., 2006], and the super- and subadditive effects of the Calvert study have not been replicated so far [for a review including the discussion of methodological problems see Hocking and Price, 2008]. In sum, the role of pSTS in higher-level audiovisual integration processes may have been overrated in the past. In addition, single cell recordings in the macaque brain have shown that regions integrating audiovisual stimuli were located in the macaque STS-STP-TPO-region [Bruce et al., 1981; Padberg et al., 2003]. In the human brain the equivalent regions are predicted to extend inferiorly from posterior STS into MTG [Van Essen, 2004] so that a distinction between functional STS and MTG or MOG activations has to be accepted only with reservation. This may be the case why some studies do not separate these regions and instead use less precise anatomical terms like “pSTS/MTG” [e.g., Beauchamp et al., 2004b].

Nevertheless, the left posterior TO junction seems to be crucially involved in the processing of speech and iconic gestures.

Processing of Unrelated Gestures With Familiar Speech

Our second hypothesis stated that for unrelated gestures mainly parietal activations would be observed, resulting from less congruency between speech and gestures.

Before discussing the fMRI results for unrelated gestures, the behavioral recognition data merit attention. Unrelated gesture recognition, as well as iconic gesture recognition, were equally enhanced compared with speech-only stimuli. This indicates that for bimodal stimuli deeper processing had occurred, most likely through offering binding opportunity. This is in line with other studies that demonstrated even better memory performances for incongruous than for congruous pictures, probably due to increased processing [Michelon et al., 2003].

On the neural level we observed bilateral activations in posterior temporal regions and in the inferior parietal lobule (IPL). In contrast to iconic gestures, the temporal activations for unrelated gestures were located more anteriorly and inferiorly and closely matched the activations observed for the [GU > GI] analysis. Besides its general involvement in semantic language processing [cf. Bookheimer, 2002; Vigneau et al., 2006], the MTG has been found activated for such aspects of object knowledge as

associations with sensorimotor correlates of their use [e.g., Chao and Martin, 2000; Martin et al., 1996]. Compared with iconic gestures, this more anterior MTG activation is interpreted as a stronger reliance on linguistic aspects, as meaning mainly could be extracted from speech, because our unrelated gestures did not contain clear semantic information.

Some authors suggest a strong link between action and language systems that could be fulfilled by the postulated mirror neuron system, including the IPL [Nishitani et al., 2005; Rizzolatti and Arbib, 1998]. Concerning the interaction of language and gestures, our results are consistent with the existing studies. Willems et al. [2007] found significant activations in the left intraparietal sulcus for the gesture mismatch condition compared with correct pairs of speech and gesture. This condition is similar to the unrelated gestures used in our study. Holle et al. [2008] did not use mismatches in their study but instead manipulated the ambiguity of their sentences and also found inferior bilateral parietal activations for gestures supporting the meaning of a sentence as opposed to grooming gestures. All of these IPL activations could be interpreted as a process of observation-execution mapping involving more simulation costs. In the study by Holle et al. [2008] the iconic gestures corresponding to the dominant and subordinate meanings were more complex than grooming and their meaning was still somewhat unclear [see Hadar and Pinchas-Zamir, 2004] and thus open to different interpretations. Thus, an initially attributed action goal may emerge as not appropriate and more simulation cycles would be needed for gestures than for grooming. For mismatching or unrelated gestures this explanation holds even more and is further supported by numerous studies showing parietal activations for the observation of meaningless hand movements [for a review see Grèzes and Decety, 2001]. Interestingly, apart from being activated by pure action observation IPL seems to be modulated by semantic information from speech. This might be accomplished by higher order cortical areas that modulate motor representations in a top-down process. But this explanation raises the question why we did not observe parietal activations for the processing of iconic gestures. In none of the difference contrasts used for the iconic conjunction analysis parietal activations were found. As our sentences were unambiguous and the gestures paralleled the speech content, it may have been that the clear language context influenced and constrained the simulation process. But even though there are some authors postulating a strong link between action and language systems [Nishitani et al., 2005; Rizzolatti and Arbib, 1998] and some studies showing influences of language domains on action processing [e.g., Bartolo et al., 2007; Gentilucci et al., 2006], this is a rather tentative explanation. In sum, the activation levels in the IPL for the processing of unrelated gestures most likely relate to an action processing component, possibly rerunning gesture information in order to create integration opportunities.

Common and Separate Systems for Iconic and Unrelated Gestures

For the semantic processing of effective and faulty stimuli usually the same networks are found. But the processing of faulty stimuli often is characterized by additional activation patterns, and this is what we have found: common to both related and unrelated speech-gesture pairs is activation in the left posterior temporal cortex. Besides this small overlap all activations retrieved by the conjunction analyses for both iconic and unrelated gestures were specific to the respective gesture category. The left middle occipital gyrus was exclusively activated for iconic gestures, whereas bilateral parietal as well as inferior and middle temporal areas were specifically observed for unrelated gestures. We interpret these distinct activation patterns as different routes of processing reflecting integration effort.

The temporal activations during the processing of speech with unrelated gestures are interpreted as a strong reliance on the linguistic network, as the meaning of the sequence can only be grasped from speech. This is in line with studies revealing activations in these regions in response to words activating object-related knowledge [e.g., Chao et al., 1999; Perani et al., 1999]. The integration process for iconic gestures as reflected in occipital activations, in contrast, seems to be based on the familiar visual features of speech contents and their representations in gestures (e.g., “spiral stair”), rendering activation of linguistic semantic areas in the temporal lobe redundant. This view is mainly supported by studies on visual imagery, showing involvement of the left middle occipital and inferior temporal gyrus in visual imagery of seen scenes [Johnson et al., 2007], walking [Jahn et al., 2004], objects like chairs in contrast to faces or houses [Ishai et al., 1999, 2000a,b] and in motion imagery of graphic stimuli [Goebel et al., 1998]. In addition, a meta-analysis of image generation in relation to lesions has noted a trend toward posterior damage [Farah, 1995]. Thus, iconic gestures possibly activate internal visual representations automatically. The observed parietal activations during the processing of unrelated speech-gesture pairs presumably indicate the gesture-oriented part of the integration process, activating the dorsal visual pathway in order to process the novel movement patterns by encoding their spatiotemporal relationships. This is substantiated by findings related to action observation (see above) and by its deconstructing function during the encoding of abstract motion sequences [Grafton et al., 1998; Jenkins et al., 1994].

There are several possible interpretations of the results of the unrelated conjunction analysis. First of all, integration of unrelated information may seem unnecessary. However, there are several studies that have shown audiovisual integration effects for unrelated or incongruent information [Hein et al., 2007; Hocking and Price, 2008; Olivetti Belardinelli et al., 2004; Taylor et al., 2006]. Second, our behavioral results suggest that effective inte-

gration has occurred because recognition of unrelated speech gesture pairs was similar to that of related pairs. Hence, it is rather unlikely that the observed activation is only the correlate of the attempt to integrate. At least parts of the revealed areas should be indicative of effective integration. Third, we cannot rule out that the regions revealed by the unrelated conjunction analysis are somewhat influenced by mismatch processing. But as both kinds of gestures commonly activated parts of the left posterior MTG, it can be assumed that even for speech with unrelated gestures the brain tries to create a common representation. We suggest that, independent of relatedness, it is the left posterior MTG where effective motion-language mapping takes place. The temporal, parietal and occipital regions found specifically activated for unrelated and iconic gestures, respectively, may serve as auxiliary regions processing single aspects of the presented stimulus in order to enable integration of these aspects in the posterior MTG. We suggest that due to the temporospatial co-occurrence of speech and gestures the brain assigns relatedness to these two sources of information. Cognitive processes try to map speech and gestures onto each other, be they related (naturally occurring) or unrelated (unnatural experimental setting). In the case of unrelated gestures, the mapping process is more difficult and requires more effort, leading to additional activations in temporal and parietal regions that probably rerun information from both modalities in order to enable integration. The core integration process in the left posterior MTG, finally, could be achieved by means of a patchy organization as it has been found in the STS [Beauchamp et al., 2004a]. With high-resolution fMRI small patches were found that responded primarily to unimodal visual or auditory input, presumably translated that information in a common code for both modalities, which then was subsequently integrated in multisensory patches that were located between the unimodal ones.

Because of the limited temporal resolution we cannot make precise inferences about the temporal sequence of cortical activity and whether these activations relate to intermediate processing stages (functions) or to later stages of accessing the end-product “representation,” i.e., the meaning of each sentence. Thus, the additional temporal and parietal activations found for the processing of unrelated gestures may either represent auxiliary processes looking for information that can be integrated or may represent additional elements of the representation of these unrelated speech gesture pairs.

Concerning the parietal activations observed for unrelated gestures alternative interpretations are possible. One might ask whether the unrelated gestures differed in kinematics from the iconic gestures. But we constructed our stimuli carefully and statistical analyses of movement extent, movement time and relations between speech and gestures did not reveal any significant differences between these two kinds of gestures (see Supporting Information Material). This suggests that it is rather unlikely that there

were any differences in kinematics that may have caused parietal activation.

The parietal activations also could stem from increased attentional processes [for a review see Culham and Kanwisher, 2001; Husain and Nachev, 2007]. Parietal regions, such as the intraparietal sulcus, are known to be involved in guiding visual or spatial attention [Corbetta and Shulman, 2002]. Though our stimuli were matched for complexity, it might be possible that the unrelated movements produced stronger attention-related parietal activations than the iconic gestures. It is possible that due to the temporospatial co-occurrence the expectation was raised that these two sources of information belong together—which then was violated by the unrelatedness of the gesture. This might have resulted in increased attention and noticeable recognition rates. But the argument of increased attention or arousal seems to apply more to paradigms where stimuli are assigned a certain task-relevance [Teder-Sälejärvi et al., 2002], something that we avoided by using a task that was neither related to the gestures nor to the language. Still, the detection of unrelatedness with subsequent increased performance rates is indicative of an integration process.

According to the study by Willems et al. [2007] one could have expected involvement of the IFG in relation to integration processes. We observed this activation only in an analysis similar to that of Willems et al., contrasting unrelated to related speech-gesture pairs [GU > GI]. In the conjunction analysis for unrelated gestures only at a more liberal explorative threshold IFG activation could be observed. For iconic gestures this was not the case. This suggests that the role of the IFG in speech gesture integration processes is not purely integrative but rather related to the detection and resolution of incompatible stimulus representations and for implementing reanalyses in the face of misinterpretations [Kuperberg et al., 2008; Novick et al., 2005]. This explanation might also account for IFG involvement in the processing of metaphoric speech-gesture pairs where the speech content cannot be taken literally (if it is taken literally there is conflict between speech and gesture) and has to be transferred to an abstract level [Kircher et al., 2009; Straube et al., in press]. This explanation is also in accordance with the one given by Willems et al. [2007], dealing with increased processing load.

Finally one might question our definition and analysis of integration-related processes. A criterion that researchers in the field of multisensory integration seem to agree upon is that for a brain region to be considered as an audiovisual integration site, it has to exhibit activation to (1) an audio-only condition, (2) a visual-only condition and (3) show additionally some property of multisensory integration like $[0 < S < SG > G > 0]$ (with S = speech, G = gestures and SG = speech plus gestures). Our data, especially the activation of the TO region, do meet this requirement for both iconic and unrelated gestures, supporting our interpretation of an integrative region.

The comparison of the two analysis approaches suggests that the conjunction approach is superior to simply con-

trasting unrelated to related gestures. Only by this means it was possible to reveal processes that are shared between related and unrelated gestures, and to delineate regions that are not better explained by mismatch detection.

CONCLUSIONS

In conclusion, our results demonstrated an area of activation shared by related iconic as well as by unrelated gestures, presumably reflecting a common integration process on a semantic level. Two distinct processing streams converge in the left posterior temporal cortex which was commonly activated by both kinds of gestures. The network for iconic gestures was located in the left temporo-occipital cortex, possibly reflecting more visually based processing. In contrast, the network for unrelated gestures was bilaterally situated in posterior temporal and inferior parietal regions, most likely reflecting two auxiliary processes split up in linguistic aspects and novel movement pattern aspects that rerun information. The key role of the left posterior temporal cortex may be due to its anatomical location between areas processing visual form and motion and on the other side visual and auditory association areas, what makes it particularly suitable for the integration of these types of information. The possibility that different subregions of this area are specialized for associating different properties within and across visual and auditory modalities (here gesture and language) remains an avenue for future exploration.

To our knowledge this is the first study revealing the neural correlates of iconic gesture processing with natural speech-gesture pairs. As even unrelated gestures presumably lead to integration processes, but in different brain areas, it is likely that other kinds of gestures (e.g., emblematic, deictic or metaphoric) are processed in specific neural networks, which are yet to be explored.

ACKNOWLEDGMENTS

The authors are grateful to all the subjects who participated in this study, to Thilo Kellermann for help with the data analysis, to Olga Sachs for linguistic assistance and to the IZKF service team for support acquiring the data.

REFERENCES

- Barsalou LW, Kyle Simmons W, Barbey AK, Wilson CD (2003): Grounding conceptual knowledge in modality-specific systems. *Trends Cogn Sci* 7:84–91.
- Bartolo A, Weisbecker A, Coello Y (2007): Linguistic and spatial information for action. *Behav Brain Res* 184:19–30.
- Baumgärtner A, Weiller C, Büchel C (2002): Event-related fMRI reveals cortical sites involved in contextual sentence integration. *Neuroimage* 16:736–745.
- Beauchamp MS (2005): Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics* 3:93–113.

- 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.
- Bruce C, Desimone R, Gross CG (1981): Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. *J Neurophysiol* 46:369–384.
- Calvert GA (2001): Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123.
- Calvert GA, Campbell R, Brammer MJ (2000): Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Curr Biol* 10:649–657.
- Chao LL, Haxby JV, Martin A (1999): Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci* 2:913–919.
- Chao LL, Martin A (2000): Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* 12:478–484.
- Corbetta M, Shulman GL (2002): Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215.
- Craigheo L, Metta G, Sandini G, Fadiga L (2007): The mirror-neurons system: Data and models. In: von Hofsten C, Rosander K, editors. *From Action to Cognition*. Amsterdam: Elsevier. pp 39–59.
- Culham JC, Kanwisher NG (2001): Neuroimaging of cognitive functions in human parietal cortex. *Curr Opin Neurobiol* 11:157–163.
- De Renzi E, Lucchelli F (1988): Ideational apraxia. *Brain* 111:1173–1185.
- Démonet JF, Thierry G, Cardebat D (2005): Renewal of the neurophysiology of language: Functional neuroimaging. *Physiol Rev* 85:49–95.
- Ebisch SJH, Babiloni C, Del Gratta C, Ferretti A, Perrucci MG, Caulo M, Sitskoorn MM, Luca Romani G (2007): Human neural systems for conceptual knowledge of proper object use: A functional magnetic resonance imaging study. *Cereb Cortex* 17:2744–2751.
- 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.
- Farah MJ (1995): Current issues in the neuropsychology of image generation. *Neuropsychologia* 33:1455–1471.
- 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.
- Gentilucci M, Bernardis P, Crisi G, Volta RD (2006): Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. *J Cogn Neurosci* 18:1059–1074.
- Goebel R, Khorram-Sefat D, Muckli L, Hacker H, Singer W (1998): The constructive nature of vision: Direct evidence from functional magnetic resonance imaging studies of apparent motion and motion imagery. *Eur J Neurosci* 10:1563–1573.
- Grafton ST, Hazeltine E, Ivry RB (1998): Abstract and effector-specific representations of motor sequences identified with PET. *J Neurosci* 18:9420–9428.
- Grèzes J, Decety J (2001): Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta analysis. *Hum Brain Mapp* 12:1–19.
- Hadar U, Krauss RK (1999): Iconic gestures: The grammatical categories of lexical affiliates. *J Neurolinguistics* 12:1–12.
- Hadar U, Pinchas-Zamir L (2004): The semantic specificity of gesture: Implications for gesture classification and function. *J Lang Soc Psychol* 23:204–214.
- Hein G, Doehrmann O, Müller 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.
- Hickok G, Poeppel D (2000): Towards a functional neuroanatomy of speech perception. *Trends Cogn Sci* 4:131–138.
- Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Hocking J, Price CJ (2008): The role of the posterior superior temporal sulcus in audiovisual processing. *Cereb Cortex* 18:2439–2449.
- Holle H, Gunter TC (2007): The role of iconic gestures in speech disambiguation: ERP evidence. *J Cogn Neurosci* 19:1175–1192.
- Holle H, Gunter TC, Rüschmeyer SA, Hennenlotter A, Iacoboni M (2008): Neural correlates of the processing of co-speech gestures. *Neuroimage* 39:2010–2024.
- Husain M, Nachev P (2007): Space and the parietal cortex. *Trends Cogn Sci* 11:30–36.
- Ishai A, Ungerleider LG, Haxby JV (2000a): Distributed neural systems for the generation of visual images. *Neuron* 28:979–990.
- Ishai A, Ungerleider LG, Martin A, Haxby JV (2000b): The representation of objects in the human occipital and temporal cortex. *J Cogn Neurosci* 12:35–51.
- Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV (1999): Distributed representation of objects in the human ventral visual pathway. *Proc Natl Acad Sci USA* 96:9379–9384.
- Jahn K, Deutschländer A, Stephan T, Strupp M, Wiesmann M, Brandt T (2004): Brain activation patterns during imagined stance and locomotion in functional magnetic resonance imaging. *Neuroimage* 22:1722–1731.
- Jenkins IH, Brooks DJ, Nixon PD, Frackowiak RS, Passingham RE (1994): Motor sequence learning: A study with positron emission tomography. *J Neurosci* 14:3775–3790.
- Johnson MR, Mitchell KJ, Raye CL, D'Esposito M, Johnson MK (2007): A brief thought can modulate activity in extrastriate visual areas: Top-down effects of refreshing just-seen visual stimuli. *Neuroimage* 37:290–299.
- 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.
- Katanoda K, Matsuda Y, Sugishita M (2002): A spatio-temporal regression model for the analysis of functional MRI data. *Neuroimage* 17:1415–1428.
- 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.
- Kita S, Ozyürek 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* 48:16–32.

- Kourtzi Z, Kanwisher N (2000): Activation in human MT/MST by static images with implied motion. *J Cogn Neurosci* 12:48–55.
- 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, McGuire PK, Bullmore ET, Brammer MJ, Rabe-Hesketh S, Wright IC, Lythgoe DJ, Williams SCR, David AS (2000): Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: An fMRI study. *J Cogn Neurosci* 12:321–341.
- 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.
- Kutas M, Hillyard SA (1980): Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207:203–205.
- Lewis JW (2006): Cortical networks related to human use of tools. *Neuroscientist* 12:211–231.
- Malikovic A, Amunts K, Schleicher A, Mohlberg H, Eickhoff SB, Wilms M, Palomero-Gallagher N, Armstrong E, Zilles K (2007): Cytoarchitectonic analysis of the human extrastriate cortex in the region of V5/MT: A probabilistic, stereotaxic map of area hOc5. *Cereb Cortex* 17:562–574.
- Martin A, Chao LL (2001): Semantic memory and the brain: Structure and processes. *Curr Opin Neurobiol* 11:194–201.
- Martin A, Wiggs CL, Ungerleider LG, Haxby JV (1996): Neural correlates of category-specific knowledge. *Nature* 379:649–652.
- McNeill D (1992): *Hand and Mind—What Gestures Reveal About Thought*. Chicago, Illinois, and London, England: The University of Chicago Press.
- Michelon P, Snyder AZ, Buckner RL, McAvoy M, Zacks JM (2003): Neural correlates of incongruous visual information: An event-related fMRI study. *Neuroimage* 19:1612–1626.
- Mummary CJ, Patterson K, Hodges JR, Price CJ (1998): Functional neuroanatomy of the semantic system: Divisible by what? *J Cogn Neurosci* 10:766–777.
- Ni W, Constable RT, Mencl WE, Pugh KR, Fulbright 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* 12:120–133.
- Nichols T, Brett M, Andersson J, Wager T, Poline JB (2005): Valid conjunction inference with the minimum statistic. *Neuroimage* 25:653–660.
- Nishitani N, Schürmann M, Amunts K, Hari R (2005): Broca's region: From action to language. *Physiology* 20:60–69.
- Novick JM, Trueswell JC, Thompson-Schill SL (2005): Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cogn Affect Behav Neurosci* 5:263–281.
- Oldfield RC (1971): The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Olivetti Belardinelli M, Sestieri C, Matteo R, Delogu F, Gratta C, Ferretti A, Caulo M, Tartaro A, Romani G (2004): Audio-visual crossmodal interactions in environmental perception: An fMRI investigation. *Cogn Process* 5:167–174.
- Ozyürek A, Willems RM, Kita S, Hagoort P (2007): On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *J Cogn Neurosci* 19:605–616.
- Padberg J, Seltzer B, Cusick CG (2003): Architectonics and cortical connections of the upper bank of the superior temporal sulcus in the rhesus monkey: An analysis in the tangential plane. *J Comp Neurol* 467:418–434.
- Perani D, Schnur T, Tettamanti M, Italy, Cappa SF, Fazio F (1999): Word and picture matching: A PET study of semantic category effects. *Neuropsychologia* 37:293–306.
- Rizzolatti G, Arbib MA (1998): Language within our grasp. *Trends Neurosci* 21:188–194.
- Rizzolatti G, Craighero L (2004): The mirror neuron system. *Ann Rev Neurosci* 27:169–192.
- Ross RS, Slotnick SD (2008): The hippocampus is preferentially associated with memory for spatial context. *J Cogn Neurosci* 20:432–446.
- Seltzer B, Pandya DN (1978): Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey. *Brain Res* 149:1–24.
- 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 (2008): Memory effects of speech and gesture binding: Cortical and hippocampal activation in relation to subsequent memory performance. *J Cogn Neurosci* [Epub ahead of print].
- Taylor KI, Moss HE, Stamatakis EA, Tyler LK (2006): Binding crossmodal object features in perirhinal cortex. *Proc Natl Acad Sci USA* 103:8239–8244.
- Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA (2002): An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. *Brain Res Cogn Brain Res* 14:106–114.
- Tranel D, Damasio H, Damasio AR (1997): A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia* 35:1319–1327.
- Tranel D, Feinstein J, Manzel K (2008): Further lesion evidence for the neural basis of conceptual knowledge for persons and other concrete entities. *J Neuropsychol* 2:301–320.
- Van Essen DC (2004): Surface-based approaches to spatial localization and registration in primate cerebral cortex. *Neuroimage* 23:S97–S107.
- Vigneau M, Beaucousin V, Hervé PY, Duffau H, Crivello F, Houdé O, Mazoyer B, Tzourio-Mazoyer N (2006): Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage* 30:1414–1432.
- Willems RM, Hagoort P (2007): Neural evidence for the interplay between language, gesture, and action: A review. *Brain Lang* 101:278–289.
- Willems RM, Ozyürek A, Hagoort P (2007): When language meets action: The neural integration of gesture and speech. *Cereb Cortex* 17:2322–2333.
- Wu YC, Coulson S (2007): How iconic gestures enhance communication: An ERP study. *Brain Lang* 101:234–245.
- Zarahn E, Aguirre GK, D'Esposito M (1997): Empirical analyses of BOLD fMRI statistics. *Neuroimage* 5:179–197.