Supramodal neural processing of abstract information conveyed by speech and gesture

Benjamin Straube[1]*, Yifei He[1,2], Miriam Steines[1], Helge Gebhardt[3], Tilo Kircher[1], Gebhard Sammer[3] and Arne Nagels[1]

[1] Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Marburg, Germany
[2] Department of General Linguistics, Johannes Gutenberg-University Mainz, Mainz, Germany
[3] Cognitive Neuroscience at Centre for Psychiatry, Justus Liebig University Giessen, Giessen, Germany

Abstractness and modality of interpersonal communication have a considerable impact on comprehension. They are relevant for determining thoughts and constituting internal models of the environment. Whereas concrete object-related information can be represented in mind irrespective of language, abstract concepts require a representation in speech. Consequently, modality-independent processing of abstract information can be expected. Here we investigated the neural correlates of abstractness (abstract vs. concrete) and modality (speech vs. gestures), to identify an abstractness-specific supramodal neural network. During fMRI data acquisition 20 participants were presented with videos of an actor either speaking sentences with an abstract-social [AS] or concrete-object-related content [CS], or performing meaningful abstract-social emblematic [AG] or concrete-object-related tool-use gestures [CG]. Gestures were accompanied by a foreign language to increase the comparability between conditions and to frame the communication context of the gesture videos. Participants performed a content judgment task referring to the person vs. object-relatedness of the utterances. The behavioral data suggest a comparable comprehension of contents communicated by speech or gesture. Furthermore, we found common neural processing for abstract information independent of modality (AS > CS ∩ AG > CG) in a left hemispheric network including the left inferior frontal gyrus (IFG), temporal pole, and medial frontal cortex. Modality specific activations were found in bilateral occipital, parietal, and temporal as well as right inferior frontal brain regions for gesture (G > S) and in left anterior temporal regions and the left angular gyrus for the processing of speech semantics (S > G). These data support the idea that abstract concepts are represented in a supramodal manner. Consequently, gestures referring to abstract concepts are processed in a predominantly left hemispheric language related neural network.

Keywords: gesture, speech, fMRI, abstract semantics, emblematic gestures, tool-use gestures

INTRODUCTION

Human communication is distinctly characterized by the ability to convey abstract concepts such as feeling, evaluations, cultural symbols, or theoretical assumptions. This can be differentiated from references to our physical environment consisting of concrete objects and their relationships to each other. In addition to our language capacity, humans also employ gestures as flexible tool to communicate both concrete and abstract information (Kita et al., 2007; Straube et al., 2011a). The investigation of abstractness and modality of communicated information can deliver important insight into the neural representation of concrete and abstract meaning. However, up to now, evidence about communalities or differences in the neural processing of abstract vs. concrete meaning communicated by speech vs. gesture is missing.

Recently, a hierarchical model of language and thought has been suggested (Perlovsky and Ilin, 2010) which proposes that abstract thinking is impossible without speech (Perlovsky and Ilin, 2013). According to this model, abstract information is processed by a neural language system, regardless of whether speech or gesture is chosen as a tool to convey this information. Following this assumption, concrete object-related information is represented in mind independent of speech and hence in a modality-dependent manner in brain regions sensitive to for example visual or motor information. The latter assumption—at least partly—contradicts existing embodiment theories, which suggest a strong overlap of the sensory-motor and language system in particular with respect to the processing of concrete concepts (Gallese and Lakoff, 2005; Arbib, 2008; Fischer and Zwaan, 2008; D'Ausilio et al., 2009; Pulvermüller and Fadiga, 2010). However, the particular role of the communication modality for the neural representation of abstract as opposed to concrete concepts has not been investigated so far.

The impact of abstractness on speech processing (e.g., Rapp et al., 2004, 2007; Eviatar and Just, 2006; Lee and Dapretto, 2006; Kircher et al., 2007; Mashal et al., 2007, 2009; Shibata et al., 2007;

Schmidt and Seger, 2009; Desai et al., 2011) and on the neural integration of speech and gesture information has been demonstrated in several functional magnetic resonance imaging (fMRI) studies using different experimental approaches (Cornejo et al., 2009; Kircher et al., 2009b; Straube et al., 2009, 2011a, 2013a; Ibáñez et al., 2011). There is converging evidence suggesting that especially the left inferior frontal gyrus (IFG) plays a decisive role in the processing of abstract semantic figurative meaning in speech (Rapp et al., 2004, 2007; Kircher et al., 2007; Shibata et al., 2007). However, results can further differ due to other factors, such as familiarity, imagibility, figurativeness, or processing difficulty (Mashal et al., 2009; Schmidt and Seger, 2009; Cardillo et al., 2010; Schmidt et al., 2010; Diaz et al., 2011).

In contrast to abstract information processing, it has been suggested that concrete information is processed in different brain regions sensitive to the specific information type: e.g., spatial information in the parietal lobe (Ungerleider and Haxby, 1994; Straube et al., 2011c), form or color information in the temporal lobe (Patterson et al., 2007). A similar finding is illustrated by Binder and Desai (2011): by reviewing 38 imaging studies that examined concrete knowledge processing during language comprehension tasks, the authors found that the processing of action-related speech material activates brain regions that are also involved in action execution (see also Hauk et al., 2004; Hauk and Pulvermüller, 2004); similarly, the processing of other concrete speech information such as sound and color all tend to show activations in areas that process these perceptual modalities (Binder and Desai, 2011). In sum, abstract information processing has been shown to recruit a mainly left-lateralized fronto-temporal neural network whereas concrete information comprehension involves rather diverse activation foci, which are primarily related to the corresponding perceptual origin.

In addition to our speech capacity, gesturing is a flexible communicative tool which humans use to communicate both concrete and abstract information via the visual modality. Previous studies on object- or person-related gesture processing have either presented pantomimes of tool or object use, hands grasping for tools or objects (e.g., Decety et al., 1997; Faillenot et al., 1997; Decety and Grèzes, 1999; Grèzes and Decety, 2001; Buxbaum et al., 2005; Filimon et al., 2007; Pierno et al., 2009; Biagi et al., 2010; Davare et al., 2010; Emmorey et al., 2010; Jastorff et al., 2010); or symbolic gestures like "thumbs up" (Nakamura et al., 2004; Molnar-Szakacs et al., 2007; Husain et al., 2009; Xu et al., 2009; Andric et al., 2013). However, few studies directly compared abstract-social (person-related) with concrete-object-related gestures. A previous study demonstrated that the left IFG is involved in the processing of expressive (emotional) in contrast to body referred and isolated (object-related) hand gestures (Lotze et al., 2006). This finding suggests that the left IFG is sensitive for the processing of abstract information irrespective of communication modality (speech or gestures).

In sum, the left IFG represents a sensitive region for abstract information processing in speech or gesture, whereas the brain areas activated by concrete information depend on communication modality and semantic content. However, whether the same neural structures are relevant for the processing of gestures and

sentences with an abstract content or gestures and sentences with a concrete content remains unknown.

Common neural networks for the processing of speech and gesture information have been suggested (Willems and Hagoort, 2007), and empirically tested in several recent studies (Xu et al., 2009; Andric and Small, 2012; Straube et al., 2012; Andric et al., 2013). Andric et al. (2013) performed an fMRI study on gesture processing presenting two different kinds of hand actions (emblematic gestures and grasping movements) and speech to their participants. Thus, either emblematic gestures—hand and arm movements conveying social or symbolic meaning (e.g., "thumbs up" for having done a good job)—or grasping movements (e.g., grasping a stapler) not carrying any semantic meaning *per se* were presented. The authors identified two different types of brain responses for the processing of emblematic gestures: the first type was related to the processing of linguistic meaning, the other type corresponded to the processing of hand actions or movements, regardless of the symbolic meaning conveyed. The latter type involved brain responses in parietal and premotor areas in connection with hand movements, whereas meaning bearing information, e.g., emblem and speech, resulted in activations in left lateral temporal and inferior frontal areas. Altogether, different modalities were involved distinguishing the level of mere perceptual recognition and interpretation of socially and culturally relevant emblematic gestures. More importantly, although lacking baseline conditions containing more concrete semantics (either in gesture or speech), the results from this study tentatively imply a common neural network for processing abstract meaning, irrespective of its input modality.

In a similar vein, Xu et al. (2009) investigated the processing of emblems and pantomimes and their corresponding speech utterances via fMRI. Their finding converges with Andric and colleagues imaging results in the sense that both input modalities activated a common, left-lateralized network encompassing inferior frontal and posterior temporal regions. However, although utilizing emblems (abstract) and pantomimes (concrete) as stimuli, the authors did not elaborate on how different levels of semantics (abstract/concrete) are processed via gesture or speech. Moreover, in a recent study from our laboratory, Straube et al. (2012) looked at less conventionalized gesture—iconic gesture, but still found a fronto-temporal network which was responsible for both the processing of gesture and speech semantics. Altogether, the three aforementioned studies unanimously suggest a common fronto-temporal neural network to be responsible for the processing of not only speech but also gesture semantics.

Although tentative proposals regarding a supramodal neural network for speech and gesture semantics have been made (Xu et al., 2009; Straube et al., 2012), it remains unclear how different levels of semantics—either concrete or abstract—are processed differently with respect to the input modalities. To date, no study results on a direct comparison between abstract and concrete semantic information processing with visual (gesture) or auditory (speech) input are available.

As hypothesized above, concrete object-related information might be represented in mind with and/or without speech, whereas abstract information could require/rely on a

representation in speech. Consequently, common processing mechanisms for the processing of speech and gesture semantics can be specifically expected when abstract (in contrast to concrete) information is communicated. Therefore, the current study focused on the neural correlates of abstractness and modality in a communication context. With a factorial manipulation of content (abstract vs. concrete) and communication modality (speech vs. gestures) we wanted to shed light on supramodal neural network properties relevant for the processing of abstract in contrast to concrete information. We tested the following alternative hypotheses: first, if only abstract concepts—activated through speech or gesture in natural communication situations—are processed in a supramodal manner, then we predict consistent neural signatures only for abstract in contrast to concrete contents across different types of communication modality. However, if concrete concepts—activated through speech or gestures—are also represented in a supramodal network, we predict overlapping neural responses for concrete in contrast to abstract contents across modality.

To manipulate abstractness and communication modality we used video clips of an actor either speaking sentences with an abstract-social [AS] or concrete-object-related content [CS], or performing meaningful abstract-social (emblematic) [AG] or concrete-object-related (tool-use) gestures [CG]. Gestures were accompanied by a foreign language (Russian) to increase the comparability between conditions and naturalness of the gesture videos where spoken language frames the communication context. We used emblematic and tool-related gestures to guarantee high comprehensibility of the gestures. During the experiments participants performed a content judgment task referring to the person vs. object-relatedness of the speech and gesture communications to ensure their attention to the semantic information and the adequate comprehension of the corresponding meaning. We hypothesized modality independent activations exclusively for the processing of abstract information (AS > CS ∩ AG > CG) in language-related regions encompassing the left inferior frontal gyrus, the left middle, and superior temporal gyrus (MTG/STG) as well as regions related to social/emotional processing such as the temporal pole, the medial frontal, and anterior cingulate cortex (ACC). In addition, modality specific activations were expected in bilateral occipital, parietal, and temporal brain regions for gesture (G > S) and in left temporal, temporo-parietal, and inferior frontal regions for the processing of speech semantics (S > G).

METHODS

PARTICIPANTS

Twenty healthy subjects (7 females) participated in the study. The mean age of the subjects was 25.4 years (*SD*: 3.42, range: 22.0–35.0). All participants were right handed (Oldfield, 1971), native German speakers and had no knowledge of Russian. All subjects had normal or corrected-to-normal vision, none reported any hearing deficits. Exclusion criteria were a history of relevant medical or psychiatric illness of the participants. All subjects gave written informed consent prior to participation in the study. The study was approved by the local ethics committee.

STIMULUS MATERIAL

Video clips were selected from a large pool of different videos. Some of them have been used in previous fMRI studies, focusing on different aspects of speech and gesture processing (Green et al., 2009; Kircher et al., 2009b; Straube et al., 2009, 2010, 2011a,b, 2012, 2013a,b; Leube et al., 2012; Mainieri et al., 2013). Here, we used emblematic and tool-related gestures and corresponding sentences to guarantee high comprehensibility of the gestures and a strong difference in abstractness between conditions. For the current analysis, 208 (26 videos per condition × 4 conditions × 2 sets) short video clips depicting an actor were used. The actor performed the following conditions: (1) German sentences with an abstract-social content [AS], (2) Russian sentences with abstract-social (emblematic) gestures [AG], (3) German sentences with a concrete-object-related content [CS], and (4) Russian sentences with concrete-object-related (tool-use) gestures [CG] (**Figure 1**). Thus, we presented videos with semantic information only in speech or only in gesture, both of them in either a highly abstract-social or a concrete-object-related version. Additionally, two bimodal meaningful speech-gesture conditions and one meaningless speech-gesture condition have been presented, which are not of interest for the current analysis.

We decided to present gestures accompanied by a foreign language to increase the comparability between conditions and the naturalness of the gesture videos where spoken language frames the communication context. All sentences had a similar grammatical structure (subject—predicate—object) and were translated into Russian for the gesture conditions. Words that sounded similar in each language were avoided. Examples for the German sentences are: "The blacksmith **hammers** on the metal plate" ("Der Schmied hämmert auf die Metallplatte"; CS condition) or "The bishop **exhorts** the believers" ("Der Bischof ermahnt die
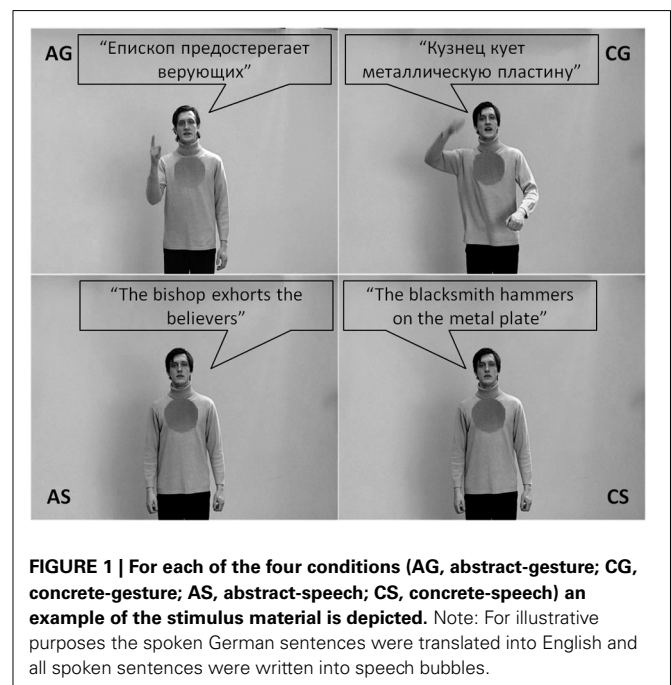

FIGURE 1 | For each of the four conditions (AG, abstract-gesture; CG, concrete-gesture; AS, abstract-speech; CS, concrete-speech) an example of the stimulus material is depicted. Note: For illustrative purposes the spoken German sentences were translated into English and all spoken sentences were written into speech bubbles.

Gläubigen"; AS condition; see **Figure 1**). Thus, the sentences had a similar length of five to eight words and a similar grammatical form, but differed considerably in content. The corresponding gestures (keyword indicated in bold) matched the corresponding speech content, but were presented here only in a foreign language context.

The same male bilingual actor (German and Russian) performed all the utterances and gestures in a natural spontaneous way. Intonation, prosody and movement characteristics in the corresponding variations of one item were closely matched. At the beginning and at the end of each clip the actor stood with arms hanging comfortably. Each clip had a duration of 5 s including 500 ms before and after the experimental manipulation, where the actor neither spoke nor moved. In the present study the semantic aspects of the stimulus material refer to differences in abstractness of the communicated information (abstract vs. concrete content).

For stimulus validation, 20 participants not taking part in the fMRI study rated each video on a scale from 1 to 7 concerning understandability, imageability and naturalness (1 = very low to 7 = very high). In order to assess *understandability* participants were asked: How understandable is the video clip? (original: "Wie VERSTÄNDLICH ist dieser Videoclip?"). The rating scale ranged from 1 = very difficult to understand (sehr schlecht verständlich) to 7 = very easy/good to understand (sehr gut verständlich). For *naturalness* ratings the participants were asked: How natural is the scene? (original: "Wie NATÜRLICH ist diese Szene?"). The rating scale ranged from 1 = very unnatural (sehr unnatürlich) to 7 = very natural (sehr natürlich). Finally, for judgments of

imageability the participants were asked: How pictorial/imageable is the scene? (original: "Wie BILDHAFT ist dieser Videoclip?"). The rating scale ranged from 1 = very abstract (sehr abstrakt) to 7 = very pictoral/imageable (sehr bildhaft). These scales have been used in previous investigations, too (Green et al., 2009; Kircher et al., 2009b; Straube et al., 2009, 2010, 2011a,b). A set of 338 video clips (52 German sentences with concrete-object-related content, 52 German sentences with abstract-social content and their counterparts in Russian-gesture and German-gesture condition and 26 Russian control condition) were chosen as stimuli for the fMRI experiment on the basis of high naturalness and high understandability for the German and gesture conditions. The stimuli were divided into two sets in order to present each participant with 182 clips during the scanning procedure (26 items per condition), counterbalanced across subjects. A single participant only saw complementary derivatives of one item, i.e., the same sentence or gesture information was only presented once per participant. This was done to avoid speech or gesture repetition or carryover effects. Again, all parameters listed above were used for an equal assignment of the video clips to the two experimental sets, to avoid set-related between-subject differences. As an overview, **Table 1** lists the mean durations of speech and gestures as well as the mean ratings of comprehension, imageability, and naturalness of the items used for the current analyses.

The ratings on understandability for the videos of the four conditions used in this study clearly show a main effect of modality, with the speech varieties scoring higher than the gesture varieties [$F_{(1, 113.51)} = 1878.79$, $P < 0.001$, two-factorial

Table 1 | Number of videos and their mean durations of stimulus parameters speech and gesture as well as their mean stimulus ratings of understandability, imageability, and naturalness according to the four conditions abstract-gesture (AG), concrete-gesture (CG), abstract-speech (AS), and concrete-speech (CS) for set 1, set 2 and in total.

| Set | Condition | N | Stimulus parameter | | | | Rating evaluations | | | | | |
| | | | Speech duration | | Gesture duration | | Understandability | | Imageability | | Naturalness | |
			Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	AG	26	2.163	0.391	2.313	0.440	3.625	0.578	4.498	0.587	4.565	0.379
	CG	26	2.303	0.434	3.033	0.364	3.537	0.808	4.785	0.695	4.340	0.540
	AS	26	2.400	0.308			6.527	0.179	3.481	0.321	4.077	0.258
	CS	26	2.332	0.290			6.650	0.209	2.967	0.308	3.181	0.293
	Total	104	2.299	0.366	2.673	0.540	5.085	1.595	3.933	0.894	4.041	0.649
2	AG	26	2.144	0.296	2.219	0.336	3.392	0.766	4.381	0.698	4.479	0.501
	CG	26	2.160	0.391	2.989	0.415	3.327	0.660	4.598	0.621	4.181	0.444
	AS	26	2.332	0.281			6.490	0.154	3.454	0.372	3.935	0.237
	CS	26	2.274	0.229			6.652	0.155	3.083	0.207	3.181	0.279
	Total	104	2.228	0.311	2.604	0.539	4.965	1.693	3.879	0.810	3.944	0.612
Total	AG	52	2.153	0.343	2.266	0.390	3.509	0.682	4.439	0.641	4.522	0.442
	CG	52	2.231	0.415	3.011	0.387	3.432	0.738	4.691	0.659	4.261	0.496
	AS	52	2.366	0.294			6.509	0.166	3.467	0.344	4.006	0.256
	CS	52	2.303	0.260			6.651	0.182	3.025	0.266	3.181	0.283
	Total	208	2.263	0.340	2.639	0.538	5.025	1.642	3.906	0.851	3.992	0.631

SD, standard deviation.

between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe]. This effect stems from the fact that different languages were used for speech only and gesture with speech conditions. Video clips with German speech scored higher than 6 while Russian speech with gestures videos scored between 3 and 4 (6.58 vs. 3.47, respectively). This difference is in line with the assumption that when presented without the respective sentence context isolated gestures are less meaningful, but even then they still are more or less understandable, which was important for the current study.

Imageability ratings indicated that there were also differences between the conditions concerning their property to evoke mental images. A significant main effect for modality showed that videos consisting of Russian sentences with gesture were evaluated as being better imaginable than videos consisting only of German sentences [4.57 vs. 3.25, respectively; $F_{(1, 144.92)} = 349.89$, $P < 0.001$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe]. A significant interaction effect indicated that this difference was even more pronounced for the concrete conditions [$F_{(1, 144.92)} = 24.22$, $P < 0.001$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe].

Naturalness ratings showed a main effect for modality as well. Videos including Russian sentences with gestures were evaluated as more natural than videos including German speech [4.39 vs. 3.59, respectively; $F_{(1, 160.63)} = 225.65$, $P < 0.001$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe]. There was also a difference in naturalness ratings concerning the abstractness of the included content. Videos depicting concrete content were evaluated as being less natural than videos depicting abstract content [4.26 vs. 3.72, respectively; $F_{(1, 160.63)} = 104.48$, $P < 0.001$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe]. Additionally, an interaction effect indicated that videos consisting of German speech with concrete content were evaluated as least natural [$F_{(1, 160.63)} = 28.18$, $P < 0.001$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe].

The sentences had an average speech duration of 2263 ms ($SD = 340$ ms), with German sentences being somewhat longer than Russian sentences [2335 vs. 2192 ms, respectively; $F_{(1, 180.94)} = 9.51$, $P < 0.05$, two-factorial between-subjects ANOVA with adjusted degrees of freedom according to Brown–Forsythe]. The gestures analyzed here had an average gesture duration of 2639 ms ($SD = 538$ ms), with gestures for concrete content being longer than gestures for abstract content [3011 vs. 2266 ms, respectively; $T_{(102)} = 9.78$, $P < 0.001$].

Events for the fMRI statistical analysis were defined in accordance with the bimodal German conditions [compare for example Green et al. (2009); Straube et al. (2012)] as the moment with the highest semantic correspondence between speech and gesture stroke (peak movement): Each sentence contained only one element that could be illustrated, which was intuitively done by the actor. The events occurred on average 2036 ms ($SD = 478$ ms) after the video start and were used for the modulation of events in the event-related fMRI analysis. The use of these predefined integration time points (see Green et al., 2009) for the fMRI data analysis had the advantage that the timing for all conditions of

one stimulus was identical since conditions were counterbalanced across subjects. Additionally, speech and gesture duration were used as parameters of no interest on single trial level to control for condition specific differences in these parameters.

EXPERIMENTAL PROCEDURE

During fMRI data acquisition participants were presented with videos of an actor either speaking sentences (S) or performing meaningful gestures (G) with an abstract-social (A) or concrete-object-related (C) content. Gestures were accompanied by an unknown foreign language (Russian). Participants performed a content judgment task referring to the person vs. object-relatedness of the utterances.

fMRI DATA ACQUISITION

All MRI data were acquired on a 3T scanner (Siemens MRT Trio series). Functional images were acquired using a T2-weighted echo planar image sequence ($TR = 2$ s, $TE = 30$ ms, flip angle 90°, slice thickness 4 mm with a 0.36 mm interslice gap, 64×64 matrix, FoV 230 mm, in-plane resolution 3.59×3.59 mm, 30 axial slices orientated parallel to the AC-PC line covering the whole brain). Two runs of 425 volumes were acquired during the experiment. The onset of each trial was synchronized to a scanner pulse.

EXPERIMENTAL DESIGN AND PROCEDURE

An experimental session comprised 182 trials (26 for each condition) and consisted of two 14-min blocks. Each block contained 91 trials with a matched number of items from each condition (13). The stimuli were presented in an event-related design in pseudo-randomized order and counterbalanced across subjects. As described above (stimulus material) across subjects each item was presented in corresponding 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. Each clip was followed by a gray background with a variable duration of 2154–5846 ms (jitter average: 4000 ms).

Before scanning, each participant received at least six practice trials outside the scanner to ensure comprehensive understanding of the experimental task. Prior to the start of the experiment, the volume of the videos was individually adjusted so that the clips were clearly audible. During scanning, participants were instructed to watch the videos and to indicate via left hand key presses whether the content of the sentence or the gesture referred to objects index finger or interpersonal social information (e.g., feelings, requests, etc.) middle finger. This task enabled us to focus participants' attention to the semantic content of speech and gesture and to investigate comprehension in a rather implicit manner. Performance rates and reaction times were recorded.

MRI DATA ANALYSIS

MR images were analyzed using Statistical Parametric Mapping (SPM8) standard routines and templates (www.fil.ion.ucl.ac.uk). After discarding the first five volumes to minimize T1-saturation effects, all images were spatially and temporally realigned, normalized (resulting voxel size $2 \times 2 \times 2$ mm^3), smoothed (8 mm isotropic Gaussian filter) and high-pass filtered (cut-off period 128 s).

Statistical whole-brain analysis was performed in a two-level, mixed-effects procedure. In the first level, single-subject BOLD responses were modeled by a design matrix comprising the onsets of each event within the videos (see stimulus material) of all seven experimental conditions. As additional factor each video phase was modeled as mini-bock with 5 s duration. To control for condition specific differences in speech and gesture duration these stimulus characteristics were used as parameters of no interest on single trial level. The hemodynamic response was modeled by the canonical hemodynamic response function (HRF). Parameter estimate (β-) images for the HRF were calculated for each condition and each subject. Parameter estimates for the four relevant conditions were entered into a within-subject flexible factorial ANOVA.

A Monte Carlo simulation of the brain volume was employed 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. Assuming an individual voxel type I error of $P < 0.001$, a cluster extent of 50 contiguous resampled voxels was indicated as necessary to correct for multiple voxel comparisons at $P < 0.05$. This cluster threshold (based on the whole brain volume) has been applied to all contrasts. The reported voxel coordinates of activation peaks are located in MNI space. For the anatomical localization, functional data were referenced to probabilistic cytoarchitectonic maps (Eickhoff et al., 2005) and the AAL toolbox (Tzourio-Mazoyer et al., 2002).

CONTRASTS OF INTEREST

The neural processing of abstract information was isolated by computing the difference contrast of abstract-social vs. concrete-object-related sentences [AS > CS] and gestures [AG > CG], whereas the opposite contrasts were applied to reveal brain regions sensitive for the processing of concrete information communicated by speech [CS > AS] and gesture [CG > AG].

In order to find regions that are commonly activated by both processes, contrasts were entered into a conjunction analysis (abstract: [AS > CS ∩ AG > CG]; concrete: [CS > AS ∩ CG > AG]), testing for independently significant effects compared at the same threshold (conjunction null, see Nichols et al., 2005).

The identical approach has been applied to demonstrate the effect of modality by calculating the following conjunctional analyses, for gesture [AG > AS ∩ CG > CS] and for speech semantics [AS > AG ∩ CS > CG].

Finally, interaction analyses were performed ([AS vs. AG] vs. [CS vs. CG]) to explore modality specific effects with regard to the processing of abstract vs. concrete information. Masking procedure has been used to ensure that all interactions are based on significant differences of the first contrast (e.g., [CG > CS] > [AG > AS] inclusively masked by [CG > CS]).

RESULTS

BEHAVIORAL RESULTS

Subjects were instructed to indicate via button press whether the actor in the video described a socially related action or an object-related action. Correct responses and their reaction times were analyzed each with a Two-Way within-subjects ANOVA with the repeated measurement factors modality (gesture vs. speech) and abstractness (abstract vs. social).

Correct responses showed a significant main effect for modality with videos depicting gesture with Russian speech receiving slightly lower scores than videos depicting German speech only [21.8 vs. 22.95 out of 26, respectively; $F_{(1, 19)} = 8.369$, $P < 0.05$, partial-eta-squared = 0.31]. A significant main effect for abstractness clearly indicated that videos describing abstract social content were less often identified correctly than videos showing concrete object-related content [20.3 vs. 24.45 out of 26, respectively; $F_{(1, 19)} = 15.361$, $P < 0.001$, partial-eta-squared = 0.45]. The factors modality and abstractness also showed a modest significant interaction effect on correct responses [$F_{(1, 19)} = 4.572$, $P < 0.05$, partial-eta-squared = 0.19] stemming from the fact that for videos depicting abstract content the difference between gesture with Russian speech and German speech was more pronounced than for videos showing concrete object-related content (**Figure 2A**).

For each participant the median reaction time for each condition was computed from all correct responses of that condition. A significant interaction effect of modality and abstractness [$F_{(1, 19)} = 5.227$, $P < 0.05$, partial-eta-squared = 0.22] indicated that while there was no difference for videos depicting concrete content, participants reacted slightly faster to videos depicting abstract content with gesture and slightly slower to videos of abstract content with German speech (**Figure 2B**).

fMRI RESULTS

Effects of modality

For the effect of gesture in contrast to speech semantics independent of the abstractness [AG > AS ∩ CG > CS] we found activation in bilateral occipital, parietal, and right frontal brain regions (see **Table 2**, and **Figure 3C**, yellow). By contrast, for the processing of speech semantics independent of abstractness [AS > AG ∩ CS > CG] we found activations in the left anterior
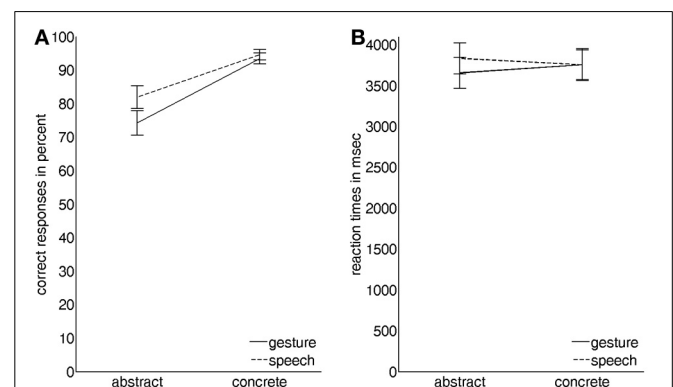

FIGURE 2 | Graphical illustration of the interaction effects of the two factors modality (gesture vs. speech) and abstractness (abstract vs. concrete) on (A) the number of correct responses in percent and on (B) the corresponding reaction times in ms (vertical lines indicate standard errors of the mean).

Table 2 | Activation peaks and anatomical regions comprising activated clusters for the conjunction contrasts representing effects of modality (speech vs. gesture and vice versa).

Contrast	Anatomical regions/hem.	No. voxels	Peak MNI coordinates			t-value
			x	y	z	
AS > AG ∩ CS > CG	Middle temporal gyrus L	673	−52	−12	−20	5.61
	Middle temporal pole L					
	Angular gyrus L	166	−54	−68	34	4.81
	Precuneus L	69	−4	−56	34	3.78
AG > AS ∩ CG > CS	Middle occipital gyrus L	6691	−48	−74	4	19.91
	Inferior temporal gyrus L					
	Middle temporal gyrus R	9536	50	−62	0	19.62
	Fusiform gyrus R					
	Superior occipital gyrus R					
	IFG, pars opercularis R	1313	44	10	28	6.72
	Middle frontal gyrus R					
	Precentral gyrus R					
	Supramarginal gyrus L	202	−62	−36	32	4.56
	Superior parietal lobe L	299	−38	−54	60	4.22
	Inferior parietal lobe L					

Table lists the respective contrast, anatomical regions, cluster size, MNI coordinates, and t-values for each significant activation (p < 0.05 corrected for multiple comparisons). MNI, Montreal Neurological Institute; AS, abstract speech; AG, abstract gesture; CS, concrete speech; CG, concrete gesture; IFG, inferior frontal gyrus; L, left; R, right.

temporal lobe and the supramarginal gyrus (see **Table 2**, and **Figure 3D**, yellow).

The exploration of general activation for each condition in contrast to low-level baseline (gray background) indicates that other regions are commonly activated in all conditions (**Figures 3A,B**). Most interestingly, the IFG seems to be activated bilaterally in the gesture conditions (**Figure 3A**) and left lateralized in the speech conditions (**Figure 3B**).

Within modality effects of abstractness

Analyses targeting at within-modality processing of abstractness in language semantics [AS > CS] showed activation in a mainly left-lateralized network encompassing an extended fronto-temporal cluster (IFG, precentral gyrus, middle, inferior, and superior temporal gyrus) as well as medial frontal regions and the right anterior middle temporal gyrus (**Table 3** and **Figure 4** top, blue). We obtained a comparable activation pattern for the within-modality processing of abstractness in gesture semantics ([AG > CG] see **Figure 4** top, yellow). The opposite contrasts revealed activation in clusters encompassing the left cerebellum, fusiform, and inferior temporal gyrus in the language contrast (CS > AS; see **Figure 4** bottom, blue) and the bilateral occipital lobe for the gesture contrast (CG > AG; see **Figure 4** bottom, yellow).

Common activations for abstractness contained in gestures and spoken language

Processing of abstract information independent of input modality as disclosed by the conjunction of [AS > CS ∩ AG > CG] was related to a left-sided frontal cluster including the temporal pole, the IFG (pars triangularis and orbitalis), the middle temporal

and angular as well as the medial superior frontal gyrus (**Table 3** and **Figure 4** top middle/right, green). The opposite conjunction analyses [CS > AS ∩ CG > AG] revealed no significant common activation for the processing of concrete in contrast to abstract information.

Interaction

No significant activation could be identified in the interaction analyses on the selected significance threshold. However, by applying a different cluster size to voxel level threshold proportion to correct for multiple comparisons (p < 0.005 and 86 voxels) as indicated by an additional Monte Carlo simulation, we found an interaction in occipital (MNI x, y, z: −20, −90, −8, t = 3.63, p < 0.001, 140 voxels), parietal (MNI x, y, z: −34, −48, 68, t = 3.80, p < 0.001, 143 voxels; MNI x, y, z: −34, −40, 48, t = 3.11, p < 0.001, 88 voxels) and premotor (MNI x, y, z: −34, −4, 62, t = 3.55, p < 0.001, 129 voxels) regions reflecting an specific increase of activation in these regions for the processing of concrete-object-related gesture meaning ([CG > CS] > [AG > AS] inclusively masked by [CG > CS]).

DISCUSSION

We hypothesized that the processing of abstract semantic information of spoken language and symbolic emblematic gestures is based on a common neural network. Our study design tailored the comparison to the level of abstract semantics, controlling for processing of general semantic meaning of speech and gesture by using highly meaningful concrete object-related information as control condition. The results demonstrate that the pathways engaged in the processing of semantics contained in both abstract spoken language and
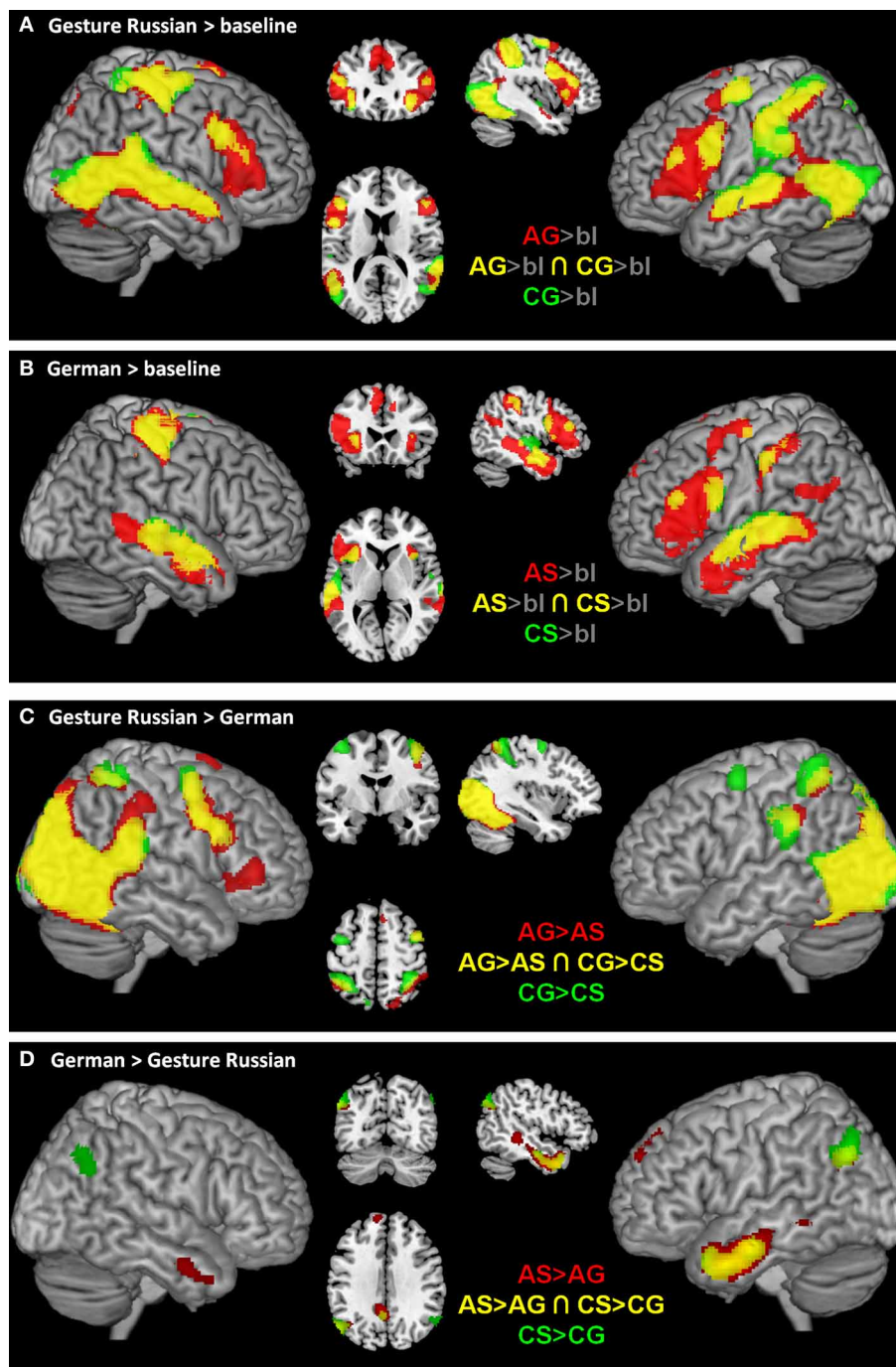
FIGURE 3 | Illustrates the fMRI results for abstract semantics (red), concrete semantics (green), and common neural structures (yellow) for each condition in contrast to low-level baseline (gray background; A, Gesture; B, German), for gesture conditions in contrast to German conditions (C) and for German in contrast to the gesture conditions (D). Results were rendered on brain slices and surface using the MRIcron toolbox (http://www.mccauslandcenter.sc.edu/mricro/mricron/install.html).

abstract-social gestures comprise the temporal pole, the IFG (pars triangularis and orbitalis), the middle temporal, angular and the superior frontal gyri. Thus, in line with our hypothesis we found modality-independent activation in a left hemispheric fronto-temporal network for the processing of abstract information. The strongly left lateralized activation pattern supports the theory that abstract semantics is independent of communication modality represented in language (at least on neural level represented in language-related brain regions).

Table 3 | Activation peaks and anatomical regions comprising activated clusters for the contrasts representing effects of abstractness (abstract vs. concrete and vice versa) dependent of modality (speech or gesture).

Contrast	Anatomical regions/hem.	No. Voxels	Peak MNI coordinates			t-value
			x	y	z	
AS > CS	Middle temporal gyrus L	3150	−52	−34	−6	5.98
	IFG, pars orbitalis L					
	Medial superior frontal gyrus L	1441	−8	56	34	5.72
	Middle temporal pole R	289	48	12	−34	5.16
	Middle temporal gyrus R					
	Angular gyrus L	458	−42	−58	24	4.41
	Precentral gyrus L	248	−38	0	62	4.36
	Precuneus L	195	−8	−50	34	4.22
AG > CG	Superior temporal pole L	910	−36	18	−24	5.43
	IFG, pars triangularis L					
	IFG, pars orbitalis L					
	Medial superior frontal gyrus L	3215	−4	30	54	5.10
	Angular gyrus L	682	−60	−60	30	4.64
	Caudate nucleus R	786	12	2	8	4.54
	Thalamus L					
	Middle temporal gyrus L	209	−48	−16	−18	4.04
AS > CS ∩ AG > CG	Medial superior frontal gyrus L	1015	−8	56	30	4.97
	Superior temporal pole L	779	−36	18	−22	4.93
	IFG, pars triangularis L					
	IFG, pars orbitalis L					
	Middle temporal gyrus L	161	−48	−14	−20	3.99
	Angular gyrus L	253	−54	−56	26	3.95
CS > AS	Cerebellum L	580	−32	−36	−28	5.95
	Inferior temporal gyrus L					
	Fusiform gyrus L					
CG > AG	Middle occipital gyrus L	1046	−44	−76	8	5.86
	Middle temporal gyrus R	285	50	−62	2	4.73
CS > AS ∩ CG > AG						n.s.

Table lists the respective contrast, anatomical regions, cluster size, MNI coordinates, and t-values for each significant activation (p < 0.05 corrected for multiple comparisons). MNI, Montreal Neurological Institute; AS, abstract speech; AG, abstract gesture; CS, concrete speech; CG, concrete gesture; IFG, inferior frontal gyrus; L, left; R, right.

EFFECTS OF MODALITY

The results of the speech [CS > CG ∩ AS > AG] and gesture contrasts [CG > CS ∩ AG > AS] clearly demonstrate that communication modality affects neural processing in the brain independent of the communication content (abstract/concrete). In line with other studies that contrasted the processing of a native against an unknown foreign language (Perani et al., 1996; Schlosser et al., 1998; Pallier et al., 2003; Straube et al., 2012), we found activation along the left temporal lobe (including STG, MTG, and ITG) for German speech contrasted with Russian speech and gesture. This strongly left-lateralized pattern has been found in all of the above mentioned studies. Apart from these studies with conditions very similar to ours, temporal as well as inferior frontal regions have been frequently implicated in various language tasks (for reviews see Bookheimer, 2002; Vigneau et al., 2006; Price, 2010). The lack of IFG activation in our study is probably dependent on the fact that we compared a native language (CS, AS) with a foreign language which was accompanied by a meaningful gesture (CG, AG). Thus, motoric or semantic processes of the left IFG might be equally involved in the speech and gesture conditions as indicated by baseline contrasts (see **Figures 3A,B**).

In line with studies on action observation (e.g., Decety et al., 1997; Decety and Grèzes, 1999; Grèzes and Decety, 2001; Filimon et al., 2007) and co-verbal gesture processing (e.g., Green et al., 2009; Kircher et al., 2009b; Straube et al., 2011a), we found for the processing of gesture in contrast to speech information a bilaterally distributed network of activation including occipital, parietal, posterior temporal, and right frontal brain regions.
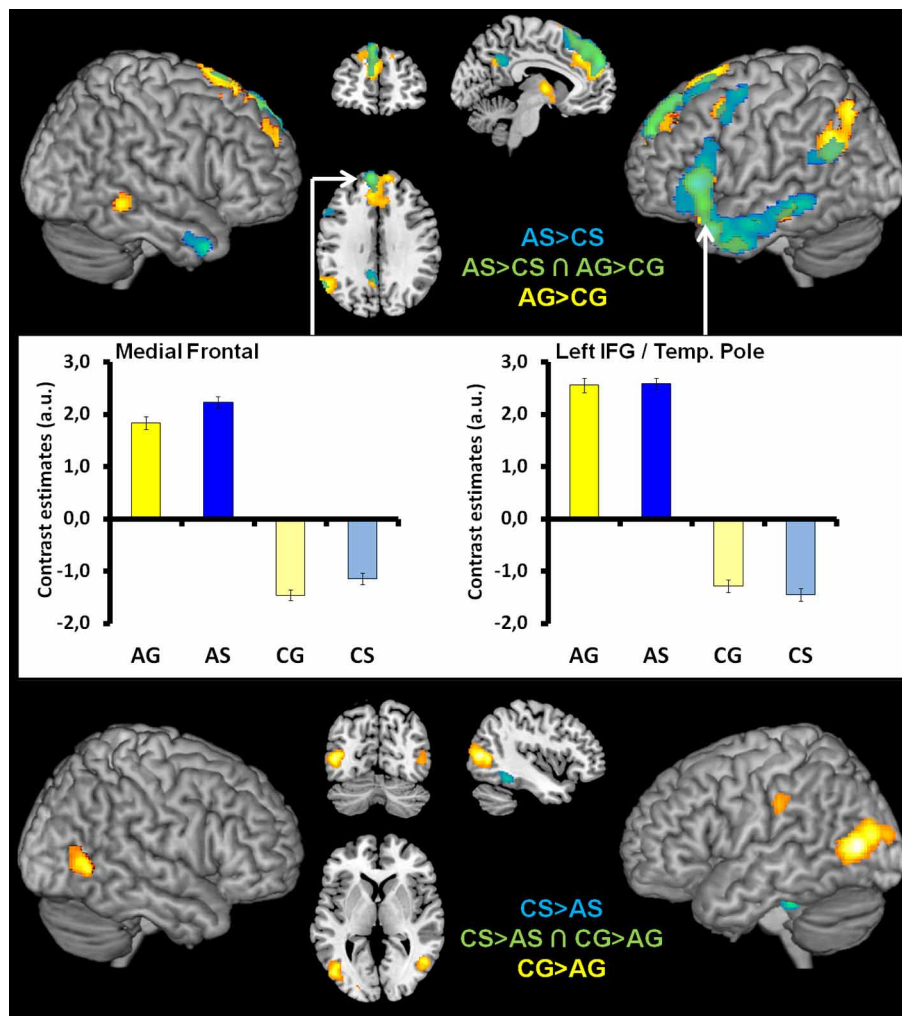
FIGURE 4 | Top illustrates the within-modality processing of abstractness in language semantics ([AS > CS], blue), gesture semantics ([AG > CG], yellow), and in common neural structures (green, overlapping regions). Bar graphs in the **middle** of figure illustrate the contrast estimates (extracted eigenvariates) for the commonly activated (green) medial superior frontal (left) and temporal pole/IFG cluster (right). These are representative for all overlapping activation clusters. The within-modality processing of concrete in contrast to abstract language semantics ([CS > AS], blue) and gesture semantics ([CG > AG], yellow) is illustrated at the **bottom** of figure. Here we found no overlap between activation patterns.

SUPRAMODAL PROCESSING OF ABSTRACT SEMANTICS OF SPEECH AND GESTURE

The processing of abstract spoken language semantics (AS > CS) and abstract semantic information conveyed through abstract-social in contrast to concrete-object-related gestures (AG > CG) activated an overlapping network of brain regions. These include a cluster in the left inferior frontal cortex (BA 44, 45) which expanded into the temporal pole, the left inferior, and middle temporal gyrus as well as a cluster in the left medial superior frontal gyrus. Those findings support the model of a supramodal semantic network for the processing of abstract information. By contrast, for concrete vs. abstract information we obtained no overlapping activation.

These results extend studies from both the gesture and the language domain (see above) in showing a common neural representation of specific speech and gesture semantics. Furthermore, the findings go beyond previous reports about common activation for symbolic gestures and speech semantics (Xu et al., 2009), in showing a specific effects for abstract but not concrete speech and gesture information. Interestingly, we previously found similar activation of the left IFG and temporal brain regions for the processing of concrete speech and gesture semantics of iconic gestures (Straube et al., 2012). Whereas iconic gestures are not symbolic and usually occur in a concrete sentence context (e.g., "The ball is round," using both hands to indicate a round shape), they might implicate rather abstract information without speech, since any concrete meaning can be revealed from these iconic

gestures in this context. Thus, the left IFG activation in our previous study could also be explained by an abstract interpretation of isolated iconic gestures (Straube et al., 2012).

The left-lateralization of our findings is congruent with the majority of fMRI studies on language (see Bookheimer, 2002; Price, 2010, for reviews). Left fronto-temporal activations have been frequently observed for semantic processing [e.g., Gaillard et al., 2004; for a review see Vigneau et al. (2006)], the decoding of meaningful actions (e.g., Decety et al., 1997; Grèzes and Decety, 2001) and also with regard to co-verbal gesture processing (Willems et al., 2007, 2009; Holle et al., 2008, 2010; Kircher et al., 2009b; Straube et al., 2011a).

With regard to the inferior frontal activations, functional imaging studies have underlined the importance of this region in the processing of language semantics. The junction of the precentral gyrus and the pars opercularis of the left IFG has been involved in controlled semantic retrieval (Thompson-Schill et al., 1997; Wiggs et al., 1999; Wagner et al., 2001), semantic priming (Sachs et al., 2008a,b, 2011; Kircher et al., 2009a; Sass et al., 2009a,b) and a supramodal network for semantic processing of words and pictures (Kircher et al., 2009a). The middle frontal gyrus (MFG) was found activated by intramodal semantic priming (e.g., Tivarus et al., 2006). However, medial frontal activation in our study might be better explained by differences in social-emotional content between conditions, which have been often found for social functioning, social cognition, theory of mind, or mentalizing (e.g., Uchiyama et al., 2006, 2012; Krach et al., 2009; Straube et al., 2010).

Since semantic memory represents the basis of semantic processing, an amodal semantic memory (Patterson et al., 2007) is a likely explanation for how speech and gesture semantics could activate a common neural network. Our findings suggest supramodal semantic processing in regions including the left temporal pole, which has been described as best candidate for a supramodal semantic "hub" (Patterson et al., 2007). Thus, abstract semantic information contained in speech and gestures might have activated supramodal semantic knowledge in our study more strongly than concrete information communicated by speech and gesture.

Our data also partially coincide with Binder and Desai's (2011) neuroanatomical model of semantic processing: in this model, low level (concrete) sensory, action and emotion semantics are processed in brain areas that are located near corresponding perceptual networks; higher-level semantics (abstract semantics), on the contrary, converges at temporal, and inferior parietal regions (Binder and Desai, 2011). Additionally, as a next step, inferior prefrontal cortices are responsible for the selection of the information stored in temporo-parietal cortices. In the current experiment, abstract information activates both temporal and inferior frontal cortices, and this could be considered as evidence supporting the role of fronto-temporal pathways in the processing of higher-level semantics. More importantly, our results suggest that this processing of abstract information is independent of input modality.

As for the processing of concrete semantics, our results are somewhat surprising because we did not find an overlap between gestural and verbal-auditory input. This result falls beyond the prediction of both strict embodiment theories (Barsalou, 1999; Gallese and Lakoff, 2005; Pulvermüller and Fadiga, 2010) and theories which propose less strict embodiment: all these theories would predict that the concrete semantics in our experiment, being predominantly action-driven, would activate motoric brain regions such as (pre-)motor and parietal cortices, and this activation pattern should be independent of the input modality. However, previous support for these theories is based on studies using single words (e.g., Willems et al., 2010; Moseley et al., 2012) instead of sentences, which might increase the task effort and specifically trigger motoring simulation. Thus, one explanation for the discrepancy between studies could be that we investigated the processing of tool-use information in a sentence context (see Tremblay and Small, 2011). Here, motoric simulation might not be necessary since contextual information facilitates semantic access (e.g., the blacksmith primes the hammer).

Our results are also in line with a recent mathematically-motivated language-cognition model proposed by Perlovsky and Ilin (2013). This model suggests that high-level abstract thinking relies on the language system and low-level and concrete thinking does not necessarily have to. Transferred to a neural perspective, both abstract meaning (irrespective of input modality) and language (processing) would recruit similar neural networks. In our experiment, the left-lateralized network for abstract meaning comprehension fits perfectly to this prediction. Although it still remains unclear how language and higher-level thinking are related at a functional level, our study provides initial neural evidence, which closely connects the two different domains.

CONCLUSION

Language is not only a communication device, but also a fundamental part of cognition and learning concepts, especially with respect to abstract concepts (Perlovsky and Ilin, 2013). In the last years the understanding of speech and gesture processing has increased; both communication channels have been disentangled and brought together again. Here we investigated the neural correlates of abstractness (abstract vs. concrete) and modality (speech vs. gestures), to demonstrate the existence of an abstractness specific supramodal neural network.

In fact, we could demonstrate the activation of a supramodal network for abstract speech and abstract gestures semantics. The identified left lateralized fronto-temporal network not only maps sound patterns and their corresponding abstract meanings in the auditory domain, but also combines gestures and their abstract meanings in the gestural-visual domain. This modality-independent network most likely gets input from modality-specific areas in the superior temporal (speech) and occipito-temporal brain regions (gestures), where the main characteristics of the spoken and gestured signals are decoded. The inferior frontal regions are responsible for the process of selection and integration, relying on more general world knowledge distributed throughout the brain (Xu et al., 2009). The challenge for future studies will be the identification of specific aspects of speech and gesture semantics or the respective format relevant for the understanding of natural receptive and productive communicative behavior and its dysfunctions in patients, for

example with schizophrenia or autism (Hubbard et al., 2012; Straube et al., 2013a,b).

REFERENCES

Andric, M., and Small, S. L. (2012). Gesture's neural language. *Front. Psychol.* 3:99. doi: 10.3389/fpsyg.2012.00099

Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., and Small, S. L. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia* 51, 1619–1629. doi: 10.1016/j.neuropsychologia.2013.03.022

Arbib, M. A. (2008). From grasp to language: embodied concepts and the challenge of abstraction. *J. Physiol. Paris* 102, 4–20. doi: 10.1016/j.jphysparis.2008.03.001

Barsalou, L. W. (1999). Perceptual symbol systems. *Behav. Brain Sci.* 22, 577–609. discussion: 610–660. doi: 10.1017/S0140525X99002149

Biagi, L., Cioni, G., Fogassi, L., Guzzetta, A., and Tosetti, M. (2010). Anterior intraparietal cortex codes complexity of observed hand movements. *Brain Res. Bull.* 81, 434–440. doi: 10.1016/j.brainresbull.2009.12.002

Binder, J. R., and Desai, R. H. (2011). The neurobiology of semantic memory. *Trends Cogn. Sci.* 15, 527–536. doi: 10.1016/j.tics.2011.10.001

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annu. Rev. Neurosci.* 25, 151–188. doi: 10.1146/annurev.neuro.25.112701.142946

Buxbaum, L. J., Kyle, K. M., and Menon, R. (2005). On beyond mirror neurons: internal representations subserving imitation and recognition of skilled object-related actions in humans. *Brain Res. Cogn. Brain Res.* 25, 226–239. doi: 10.1016/j.cogbrainres.2005.05.014

Cardillo, E. R., Schmidt, G. L., Kranjec, A., and Chatterjee, A. (2010). Stimulus design is an obstacle course: 560 matched literal and metaphorical sentences for testing neural hypotheses about metaphor. *Behav. Res. Methods* 42, 651–664. doi: 10.3758/BRM.42.3.651

Cornejo, C., Simonetti, F., Ibáñez, A., Aldunate, N., Ceric, F., López, V., et al. (2009). Gesture and metaphor comprehension: electrophysiological evidence of cross-modal coordination by audiovisual stimulation. *Brain Cogn.* 70, 42–52. doi: 10.1016/j.bandc.2008.12.005

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., and Fadiga, L. (2009). The motor somatotopy of speech perception. *Curr. Biol.* 19, 381–385. doi: 10.1016/j.cub.2009.01.017

Davare, M., Rothwell, J. C., and Lemon, R. N. (2010). Causal connectivity between the human anterior intraparietal area and premotor cortex during grasp. *Curr. Biol.* 20, 176–181. doi: 10.1016/j.cub.2009.11.063

Decety, J., and Grèzes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends Cogn. Sci.* 3, 172–178. doi: 10.1016/S1364-6613(99)01312-1

Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., et al. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain* 120(Pt 10), 1763–1777. doi: 10.1093/brain/120.10.1763

Desai, R. H., Binder, J. R., Conant, L. L., Mano, Q. R., and Seidenberg, M. S. (2011). The neural career of sensory-motor metaphors. *J. Cogn. Neurosci.* 23, 2376–2386. doi: 10.1162/jocn.2010.21596

Diaz, M. T., Barrett, K. T., and Hogstrom, L. J. (2011). The influence of sentence novelty and figurativeness on brain activity. *Neuropsychologia* 49, 320–330. doi: 10.1016/j.neuropsychologia.2010.12.004

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., et al. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* 25, 1325–1335. doi: 10.1016/j.neuroimage.2004.12.034

Emmorey, K., Xu, J., Gannon, P., Goldin-Meadow, S., and Braun, A. (2010). CNS activation and regional connectivity during pantomime observation: no engagement of the mirror neuron system for deaf signers. *Neuroimage* 49, 994–1005. doi: 10.1016/j.neuroimage.2009.08.001

Eviatar, Z., and Just, M. A. (2006). Brain correlates of discourse processing: an fMRI investigation of irony and conventional metaphor comprehension. *Neuropsychologia* 44, 2348–2359. doi: 10.1016/j.neuropsychologia.2006.05.007

Faillenot, I., Toni, I., Decety, J., Grégoire, M. C., and Jeannerod, M. (1997). Visual pathways for object-oriented action and object recognition: functional anatomy with PET. *Cereb. Cortex* 7, 77–85. doi: 10.1093/cercor/7.1.77

Filimon, F., Nelson, J. D., Hagler, D. J., and Sereno, M. I. (2007). Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *Neuroimage* 37, 1315–1328. doi: 10.1016/j.neuroimage.2007.06.008

Fischer, M. H., and Zwaan, R. A. (2008). Embodied language: a review of the role of the motor system in language comprehension. *Q. J. Exp. Psychol. (Hove)* 61, 825–850. doi: 10.1080/17470210701623605

Gaillard, W. D., Balsamo, L., Xu, B., McKinney, C., Papero, P. H., Weinstein, S., et al. (2004). fMRI language task panel improves determination of language dominance. *Neurology* 63, 1403–1408. doi: 10.1212/01.WNL.0000141852.65175.A7

Gallese, V., and Lakoff, G. (2005). The brain's concepts: the role of the Sensory-motor system in conceptual knowledge. *Cogn. Neuropsychol.* 22, 455–479. doi: 10.1080/02643290442000310

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., et al. (2009). Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Hum. Brain Mapp.* 30, 3309–3324. doi: 10.1002/hbm.20753

Grèzes, J., and Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* 12, 1–19. doi: 10.1002/1097-0193(200101)12:1<1::AID-HBM10>3.0.CO;2-V

Hauk, O., Johnsrude, I., and Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron* 41, 301–307. doi: 10.1016/S0896-6273(03)00838-9

Hauk, O., and Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex. *Hum. Brain Mapp.* 21, 191–201. doi: 10.1002/hbm.10157

Holle, H., Gunter, T. C., Ruschemeyer, S. A., Hennenlotter, A., and Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage* 39, 2010–2024. doi: 10.1016/j.neuroimage.2007.10.055

Holle, H., Obleser, J., Rueschemeyer, S. A., and Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage* 49, 875–884. doi: 10.1016/j.neuroimage.2009.08.058

Hubbard, A. L., McNealy, K., Scott-Van Zeeland, A. A., Callan, D. E., Bookheimer, S. Y., and Dapretto, M. (2012). Altered integration of speech and gesture in children with autism spectrum disorders. *Brain Behav.* 2, 606–619. doi: 10.1002/brb3.81

Husain, F. T., Patkin, D. J., Thai-Van, H., Braun, A. R., and Horwitz, B. (2009). Distinguishing the processing of gestures from signs in deaf individuals: an fMRI study. *Brain Res.* 1276, 140–150. doi: 10.1016/j.brainres.2009.04.034

Ibáñez, A., Toro, P., Cornejo, C., Urquina, H., Hurquina, H., Manes, F., et al. (2011). High contextual sensitivity of metaphorical expressions and gesture blending: a video event-related potential design. *Psychiatry Res.* 191, 68–75. doi: 10.1016/j.pscychresns.2010.08.008

Jastorff, J., Begliomini, C., Fabbri-Destro, M., Rizzolatti, G., and Orban, G. A. (2010). Coding observed motor acts: different organizational principles in the parietal and premotor cortex of humans. *J. Neurophysiol.* 104, 128–140. doi: 10.1152/jn.00254.2010

Kircher, T., Sass, K., Sachs, O., and Krach, S. (2009a). Priming words with pictures: neural correlates of semantic associations in a cross-modal priming task using fMRI. *Hum. Brain Mapp.* 30, 4116–4128. doi: 10.1002/hbm.20833

Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., et al. (2009b). Neural interaction of speech and gesture: differential activations of metaphoric coverbal gestures. *Neuropsychologia* 47, 169–179. doi: 10.1016/j.neuropsychologia.2008.08.009

Kircher, T. T., Leube, D. T., Erb, M., Grodd, W., and Rapp, A. M. (2007). Neural correlates of metaphor processing in schizophrenia. *Neuroimage* 34, 281–289. doi: 10.1016/j.neuroimage.2006.08.044

Kita, S., de Condappa, O., and Mohr, C. (2007). Metaphor explanation attenuates the right-hand preference for depictive co-speech gestures that imitate actions. *Brain Lang.* 101, 185–197. doi: 10.1016/j.bandl.2006.11.006

Krach, S., Blumel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., et al. (2009). Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. *BMC Neurosci.* 10:9. doi: 10.1186/1471-2202-10-9

Lee, S. S., and Dapretto, M. (2006). Metaphorical vs. literal word meanings: fMRI evidence against a selective role of the right hemisphere. *Neuroimage* 29, 536–544. doi: 10.1016/j.neuroimage.2005.08.003

Leube, D., Straube, B., Green, A., Blümel, I., Prinz, S., Schlotterbeck, P., et al. (2012). A possible brain network for representation of cooperative behavior and its implications for the psychopathology of schizophrenia. *Neuropsychobiology* 66, 24–32. doi: 10.1159/000337131

Lotze, M., Heymans, U., Birbaumer, N., Veit, R., Erb, M., Flor, H., et al. (2006). Differential cerebral activation during observation of expressive gestures and motor acts. *Neuropsychologia* 44, 1787–1795. doi: 10.1016/j.neuropsychologia.2006.03.016

Mainieri, A. G., Heim, S., Straube, B., Binkofski, F., and Kircher, T. (2013). Differential role of mentalizing and the mirror neuron system in the imitation of communicative gestures. *NeuroImage* 81, 294–305. doi: 10.1016/j.neuroimage.2013.05.021

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M. (2007). An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain Lang.* 100, 115–126. doi: 10.1016/j.bandl.2005.10.005

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M. (2009). An fMRI study of processing novel metaphoric sentences. *Laterality* 14, 30–54. doi: 10.1080/13576500802049433

Molnar-Szakacs, I., Wu, A. D., Robles, F. J., and Iacoboni, M. (2007). Do you see what I mean? Corticospinal excitability during observation of culture-specific gestures. *PLoS ONE* 2:e626. doi: 10.1371/journal.pone.0000626

Moseley, R., Carota, F., Hauk, O., Mohr, B., and Pulvermüller, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cereb. Cortex* 22, 1634–1647. doi: 10.1093/cercor/bhr238

Nakamura, A., Maess, B., Knösche, T. R., Gunter, T. C., Bach, P., and Friederici, A. D. (2004). Cooperation of different neuronal systems during hand sign recognition. *Neuroimage* 23, 25–34. doi: 10.1016/j.neuroimage.2004.04.034

Nichols, T., Brett, M., Andersson, J., Wager, T., and Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage* 25, 653–660. doi: 10.1016/j.neuroimage.2004.12.005

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113. doi: 10.1016/0028-3932(71)90067-4

Pallier, C., Dehaene, S., Poline, J. B., LeBihan, D., Argenti, A. M., Dupoux, E., et al. (2003). Brain imaging of language plasticity in adopted adults: can a second language replace the first? *Cereb. Cortex* 13, 155–161. doi: 10.1093/cercor/13.2.155

Patterson, K., Nestor, P. J., and Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci* 8, 976–987. doi: 10.1038/nrn2277

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S. F., Dupoux, E., et al. (1996). Brain processing of native and foreign languages. *Neuroreport* 7, 2439–2444. doi: 10.1097/00001756-199611040-00007

Perlovsky, L. I., and Ilin, R. (2010). Neurally and mathematically motivated architecture for language and thought. *Open Neuroimag. J.* 4, 70–80. doi: 10.2174/1874440001004010070

Perlovsky, L. I., and Ilin, R. (2013). Mirror neurons, language, and embodied cognition. *Neural Netw.* 41, 15–22. doi: 10.1016/j.neunet.2013.01.003

Pierno, A. C., Tubaldi, F., Turella, L., Grossi, P., Barachino, L., Gallo, P., et al. (2009). Neurofunctional modulation of brain regions by the observation of pointing and grasping actions. *Cereb. Cortex* 19, 367–374. doi: 10.1093/cercor/bhn089

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in (2009). *Ann. N. Y. Acad. Sci.* 1191, 62–88. doi: 10.1111/j.1749-6632.2010.05444.x

Pulvermüller, F., and Fadiga, L. (2010). Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11, 351–360. doi: 10.1038/nrn2811

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., and Kircher, T. T. (2004). Neural correlates of metaphor processing. *Brain Res. Cogn. Brain Res.* 20, 395–402. doi: 10.1016/j.cogbrainres.2004.03.017

Rapp, A. M., Leube, D. T., Erb, M., Grodd, W., and Kircher, T. T. (2007). Laterality in metaphor processing: lack of evidence from functional magnetic resonance imaging for the right hemisphere theory. *Brain Lang.* 100, 142–149. doi: 10.1016/j.bandl.2006.04.004

Sachs, O., Weis, S., Krings, T., Huber, W., and Kircher, T. (2008a). Categorical and thematic knowledge representation in the brain: neural correlates of taxonomic and thematic conceptual relations. *Neuropsychologia* 46, 409–418. doi: 10.1016/j.neuropsychologia.2007.08.015

Sachs, O., Weis, S., Zellagui, N., Huber, W., Zvyagintsev, M., Mathiak, K., et al. (2008b). Automatic processing of semantic relations in fMRI: neural activation during semantic priming of taxonomic and thematic categories. *Brain Res.* 1218, 194–205. doi: 10.1016/j.brainres.2008.03.045

Sachs, O., Weis, S., Zellagui, N., Sass, K., Huber, W., Zvyagintsev, M., et al. (2011). How different types of conceptual relations modulate brain activation during semantic priming. *J. Cogn. Neurosci.* 23, 1263–1273. doi: 10.1162/jocn.2010.21483

Sass, K., Krach, S., Sachs, O., and Kircher, T. (2009a). Lion - tiger - stripes: neural correlates of indirect semantic priming across processing modalities. *Neuroimage* 45, 224–236. doi: 10.1016/j.neuroimage.2008.10.014

Sass, K., Sachs, O., Krach, S., and Kircher, T. (2009b). Taxonomic and thematic categories: neural correlates of categorization in an auditory-to-visual priming task using fMRI. *Brain Res.* 1270, 78–87. doi: 10.1016/j.brainres.2009.03.013

Schlosser, M. J., Aoyagi, N., Fulbright, R. K., Gore, J. C., and McCarthy, G. (1998). Functional MRI studies of auditory comprehension. *Hum. Brain Mapp.* 6, 1–13. doi: 10.1002/(SICI)1097-0193(1998)6:1<1::AID-HBM1>3.0.CO;2-7

Schmidt, G. L., Kranjec, A., Cardillo, E. R., and Chatterjee, A. (2010). Beyond laterality: a critical assessment of research on the neural basis of metaphor. *J. Int. Neuropsychol. Soc.* 16, 1–5. doi: 10.1017/S1355617709990543

Schmidt, G. L., and Seger, C. A. (2009). Neural correlates of metaphor processing: the roles of figurativeness, familiarity and difficulty. *Brain Cogn.* 71, 375–386. doi: 10.1016/j.bandc.2009.06.001

Shibata, M., Abe, J., Terao, A., and Miyamoto, T. (2007). Neural mechanisms involved in the comprehension of metaphoric and literal sentences: an fMRI study. *Brain Res.* 1166, 92–102. doi: 10.1016/j.brainres.2007.06.040

Slotnick, S. D., and Schacter, D. L. (2004). A sensory signature that distinguishes true from false memories. *Nat. Neurosci.* 7, 664–672. doi: 10.1038/nn1252

Straube, B., Green, A., Bromberger, B., and Kircher, T. (2011a). The differentiation of iconic and metaphoric gestures: common and unique integration processes. *Hum. Brain Mapp.* 32, 520–533. doi: 10.1002/hbm.21041

Straube, B., Green, A., Chatterjee, A., and Kircher, T. (2011b). Encoding social interactions: the neural correlates of true and false memories. *J. Cogn. Neurosci.* 23, 306–324. doi: 10.1162/jocn.2010.21505

Straube, B., Wolk, D., and Chatterjee, A. (2011c). The role of the right parietal lobe in the perception of causality: a tDCS study. *Exp. Brain Res.* 215, 315–325. doi: 10.1007/s00221-011-2899-1

Straube, B., Green, A., Jansen, A., Chatterjee, A., and Kircher, T. (2010). Social cues, mentalizing and the neural processing of speech accompanied by gestures. *Neuropsychologia* 48, 382–393. doi: 10.1016/j.neuropsychologia.2009.09.025

Straube, B., Green, A., Sass, K., Kirner-Veselinovic, A., and Kircher, T. (2013a). Neural integration of

speech and gesture in schizophrenia: evidence for differential processing of metaphoric gestures. *Hum. Brain Mapp.* 34, 1696–1712. doi: 10.1002/hbm.22015

Straube, B., Green, A., Sass, K., and Kircher, T. (2013b). Superior temporal sulcus disconnectivity during processing of metaphoric gestures in schizophrenia. *Schizophr. Bull.* doi: 10.1093/schbul/sbt110. [Epub ahead of print]. Available online at: http://schizophreniabulletin.oxfordjournals.org/content/early/2013/08/16/schbul.sbt110.full.pdf

Straube, B., Green, A., Weis, S., Chatterjee, A., and 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. doi: 10.1162/jocn.2009.21053

Straube, B., Green, A., Weis, S., and Kircher, T. (2012). A supramodal neural network for speech and gesture semantics: an fMRI study. *PLoS ONE* 7:e51207. doi: 10.1371/journal.pone.0051207

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., and Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: a reevaluation. *Proc. Natl. Acad. Sci. U.S.A.* 94, 14792–14797. doi: 10.1073/pnas.94.26.14792

Tivarus, M. E., Ibinson, J. W., Hillier, A., Schmalbrock, P., and Beversdorf, D. Q. (2006). An fMRI study of semantic priming: modulation of brain activity by varying semantic distances. *Cogn. Behav. Neurol.* 19, 194–201.

Tremblay, P., and Small, S. L. (2011). From language comprehension to action understanding and back again. *Cereb. Cortex* 21, 1166–1177. doi: 10.1093/cercor/bhq189

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Uchiyama, H., Seki, A., Kageyama, H., Saito, D. N., Koeda, T., Ohno, K., et al. (2006). Neural substrates of sarcasm: a functional magnetic-resonance imaging study. *Brain Res.* 1124, 100–110. doi: 10.1016/j.brainres.2006.09.088

Uchiyama, H. T., Saito, D. N., Tanabe, H. C., Harada, T., Seki, A., Ohno, K., et al. (2012). Distinction between the literal and intended meanings of sentences: a functional magnetic resonance imaging study of metaphor and sarcasm. *Cortex* 48, 563–583. doi: 10.1016/j.cortex.2011.01.004

Ungerleider, L. G., and Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr. Opin. Neurobiol.* 4, 157–165. doi: 10.1016/0959-4388(94)90066-3

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., et al. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage* 30, 1414–1432. doi: 10.1016/j.neuroimage.2005.11.002

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., and Poldrack, R. A. (2001). Recovering meaning: left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31, 329–338. doi: 10.1016/S0896-6273(01)00359-2

Wiggs, C. L., Weisberg, J., and Martin, A. (1999). Neural correlates of semantic and episodic memory retrieval. *Neuropsychologia* 37, 103–118. doi: 10.1016/S0028-3932(98)00044-X

Willems, R. M., and Hagoort, P. (2007). Neural evidence for the interplay between language, gesture, and action: a review. *Brain Lang.* 101, 278–289. doi: 10.1016/j.bandl.2007.03.004

Willems, R. M., Hagoort, P., and Casasanto, D. (2010). Body-specific representations of action verbs: neural evidence from right- and left-handers. *Psychol. Sci.* 21, 67–74. doi: 10.1177/0956797609354072

Willems, R. M., Ozyurek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17, 2322–2333. doi: 10.1093/cercor/bhl141

Willems, R. M., Ozyurek, A., and Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage* 47, 1992–2004. doi: 10.1016/j.neuroimage.2009.05.066

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., and Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proc. Natl. Acad. Sci. U.S.A.* 106, 20664–20669. doi: 10.1073/pnas.0909197106