

# Co-Speech Gestures Influence Neural Activity in Brain Regions Associated With Processing Semantic Information

Anthony Steven Dick,<sup>1</sup> Susan Goldin-Meadow,<sup>2</sup> Uri Hasson,<sup>3</sup>  
Jeremy I. Skipper,<sup>4</sup> and Steven L. Small<sup>1,2,\*</sup>

<sup>1</sup>Department of Neurology, The University of Chicago, Chicago, Illinois

<sup>2</sup>Department of Psychology, The University of Chicago, Chicago, Illinois

<sup>3</sup>Center for Mind/Brain Sciences and Faculty of Cognitive Sciences, The University of Trento, Trento, Italy

<sup>4</sup>Sackler Institute for Developmental Psychobiology, Weill Medical College of Cornell University, New York



**Abstract:** Everyday communication is accompanied by visual information from several sources, including co-speech gestures, which provide semantic information listeners use to help disambiguate the speaker's message. Using fMRI, we examined how gestures influence neural activity in brain regions associated with processing semantic information. The BOLD response was recorded while participants listened to stories under three audiovisual conditions and one auditory-only (speech alone) condition. In the first audiovisual condition, the storyteller produced gestures that naturally accompany speech. In the second, the storyteller made semantically unrelated hand movements. In the third, the storyteller kept her hands still. In addition to inferior parietal and posterior superior and middle temporal regions, bilateral posterior superior temporal sulcus and left anterior inferior frontal gyrus responded more strongly to speech when it was further accompanied by gesture, regardless of the semantic relation to speech. However, the right inferior frontal gyrus was sensitive to the semantic import of the hand movements, demonstrating more activity when hand movements were semantically unrelated to the accompanying speech. These findings show that perceiving hand movements during speech modulates the distributed pattern of neural activation involved in both biological motion perception and discourse comprehension, suggesting listeners attempt to find meaning, not only in the words speakers produce, but also in the hand movements that accompany speech. *Hum Brain Mapp* 00:000–000, 2009. © 2009 Wiley-Liss, Inc.

**Key words:** discourse comprehension; fMRI; gestures; semantic processing; inferior frontal gyrus



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

Contract grant sponsor: National Institutes of Health; Contract grant numbers: P01HD040605, R01DC003378, F32DC008909.

\*Correspondence to: Steven L. Small, Human Neuroscience Laboratory, Department of Neurology, The University of Chicago, 5841 S. Maryland Ave. MC-2030, Chicago, IL 60637. E-mail: small@uchicago.edu

Received for publication 27 August 2008; Revised 10 February 2009; Accepted 11 February 2009

DOI: 10.1002/hbm.20774

Published online in Wiley InterScience ([www.interscience.wiley.com](http://www.interscience.wiley.com)).

## INTRODUCTION

Face-to-face communication is based on more than speech alone. Audible speech is only one component of a communication system that also includes co-speech gestures—hand and arm movements that accompany spoken language [Kendon, 1994; McNeill, 1992, 2005]. Such co-speech gestures serve an important role in face-to-face communication for both speaker and listener. Listeners not only process the words that speakers produce, but also continuously integrate gestures with speech and with other visual information (e.g., the speaker's lips, mouth, and eyes) to arrive at the speaker's meaning [Goldin-Meadow, 2006; Kendon, 1994; McNeill, 2005]. Despite the

importance of co-speech gesture to communicative interactions, little work has explored the neurobiology of how the visual information conveyed in gesture is integrated with speech. In this article, we investigate how the brain processes meaningful, as well as non-meaningful, hand movements during comprehension of speech. Our main premise is to understand the concepts conveyed by co-speech gesture, the listener's brain must (a) process biologically relevant information about the speaker's movements and, conjointly, (b) ascertain the meanings of these movements and relate them to the information conveyed in speech. In naturalistic audiovisual speech, the body is in constant motion, and actions can be parsed as potentially relevant or irrelevant to the speaker's spoken message. Different brain regions likely serve as "integration zones" for gesture and speech, but it remains unclear which process the biologically relevant aspect of gesture's motion form, and which connect gesture's semantic information with speech. Prior work suggests premotor cortex and more posterior cortical regions (i.e., posterior superior temporal gyrus [STGp] and sulcus [STS], posterior middle temporal gyrus [MTGp] and angular and supramarginal gyrus of the inferior parietal lobule [IPL]) are involved in comprehension of action, and that inferior frontal gyrus (IFG) is involved in top-down integration or determination of actions as potentially relevant or irrelevant to the speech it accompanies. This study investigates the role of these brain regions, and in particular attempts to clarify the role of STSp and IFG in connecting meaningful co-speech gestures with auditory discourse.

### **Co-speech Gesture and Action Comprehension**

As hand movements that are potentially related to speech, all gesture types must at some level be processed as motion, but while in the most general sense *gesture* refers to a category of motion, there are several types of gesture that differ in how they relate semantically to the speech they accompany [Kendon, 1988; McNeill, 2005]. For example, "beat" gestures emphasize rhythmic aspects of speech, while conventionalized symbols called "emblems" [for example "thumbs up"; McNeill, 2005] contribute meaning independent of speech. Iconic gestures, on the other hand, are meaningful gestures that can be interpreted only in the context of speech. For example, wiggling fingers up and down in the air can enrich the sentence, "He is working on the paper," allowing the listener to identify the movement as a gesture for typing. However, when the same movement accompanies the sentence, "The spider moved its legs," the listener will identify the movement as a gesture for a spider [McNeill, 2005].

All gesture types require processing biologically relevant motion, and a broad literature in both monkeys [Oram and Perrett, 1994, 1996] and humans has identified cortical regions sensitive to this type of motion—including partic-

ular STSp, but also STGp, MTGp, IPL [Beauchamp et al., 2003; Bonda et al., 1996; Grezes et al., 2001; Grossman et al., 2000; Peelen et al., 2006; Puce and Perrett, 2003; Saygin et al., 2004; Vaina et al., 2001] and premotor cortex [Saygin et al., 2004]. Notably, STSp is also thought to be an important site for integrating auditory and visual information during speech comprehension [Calvert, 2001; Calvert et al., 2000; Miller and D'Esposito, 2005; Olson et al., 2002; Sekiyama et al., 2003; Wright et al., 2003]. Consistent with these findings, recent work on gesture has found that these brain regions involved in action comprehension (e.g., premotor cortex, IPL, and STSp and STGp) respond more strongly to gesture accompanied by speech than to speech alone [Holle et al., 2008; Hubbard et al., 2009; Kircher et al., 2009; Willems et al., 2007; Wilson et al., 2008].

### **Co-speech Gesture and Semantic Processing**

Although some have argued that co-speech gestures play no role for the listener [Krauss et al., 1995], considerable behavioral research suggests they do make an important contribution to language comprehension [Beattie and Shovelton, 1999; Cohen and Otterbein, 1992; Feyereisen, 2006; Goldin-Meadow et al., 1999; Goldin-Meadow and Singer, 2003; Graham and Heywood, 1975; Kelly et al., 1999; McNeil et al., 2000; McNeill, 1992; Riseborough, 1981; Thompson and Massaro, 1994]. Event-related potential (ERP) studies provide further suggestion that gesture contributes accessible semantic information during language comprehension; the well-characterized N400 deflection, thought to reflect semantic integration [Kutas and Federmeier, 2000; Kutas and Hillyard, 1980], is sensitive to the relationship between gesture and speech [Holle and Gunter, 2007; Kelly et al., 2004, 2007; Özyürek et al., 2007; Wu and Coulson, 2005, 2007a,b]. The cortical source of the N400 deflection has, however, been difficult to ascertain [Van Petten and Luka, 2006 for review]. Some authors claim a distributed left-lateralized temporal and parietal network provides the cortical source for the N400, with little to no frontal involvement [Friederici et al., 1998, 1999; Hagoort et al., 1996; Halgren et al., 2002; Helenius et al., 2002; Kwon et al., 2005; Simos et al., 1997; Swaab et al., 1997], while others point to IFG, particularly on the left [Hagoort et al., 2004; also see Table I in Van Petten and Luka, 2006]. Along similar lines, recent investigations of co-speech gesture have failed to determine which brain regions are involved in processing the semantic aspects of gesture; some authors have emphasized the role of IFG in processing the semantic contributions of gesture [Skipper et al., 2007a; Straube et al., 2008; Willems et al., 2007], whereas others have suggested STSp is an important locus of gesture-speech integration [Holle et al., 2008].

Investigations using the same experimental paradigm in ERP and fMRI have the potential to be particularly informative with respect to this question because they can establish an N400 ERP effect of gesture, and, at the same time,

locate a cortical source for the effect. Unfortunately, two recent investigations of gesture using both ERP and fMRI provide contradictory findings. Using fMRI, Willems et al. [2007] replicated their prior ERP study [Özyürek et al., 2007]—words and gestures that were incongruent with a sentence context elicited a stronger N400 effect than gestures that were congruent. They found that left anterior IFG (*pars triangularis*; IFGTr), as well as premotor cortex, was more active when either the gesture or the word was incompatible with the sentence context (e.g., IFGTr responded more strongly when the sentence, “He should not forget the items that he *wrote* on the shopping list,” was accompanied by the incompatible iconic gesture “hit” when the verb “wrote” was spoken). They argued that greater activation in this region reflects the additional activity required to integrate the incompatible gesture with speech.

In an ERP study, Holle and Gunter [2007] found that subordinate gestures or meaningless hand movements accompanying an ambiguous homonym elicited a stronger N400 effect than dominant gestures, and they used fMRI to examine the cortical source of these effects [Holle et al., 2008]. To illustrate, the sentence “She touched the mouse” could be accompanied by a gesture matching the dominant meaning (e.g., miming holding a live mouse by the tail), the subordinate meaning (miming grasping a computer mouse), or by a non-related grooming movement (scratching the chin). Contrary to the findings reported by Willems et al. [2007], Holle et al. failed to find any left IFG involvement in integrating gesture with speech. Instead, left STSp responded more strongly to iconic gestures than meaningless grooming movements, suggesting this region (but not left IFG) was sensitive to the content of the gestures.

Other functional imaging studies of gesture have provided mixed evidence for the importance of left IFG in processing the semantic aspects of gesture. Although Kircher et al. [2009] found left IFG was more active for audiovisual speech with gesture compared to auditory-only speech, Wilson et al. [2008] found no differences in IFG for this comparison. However, Straube et al. [2008] found subsequent memory for meaningful metaphoric gestures and non-meaningful hand movements accompanying sentences was positively correlated with bilateral IFG activation (although right IFG was more sensitive to non-meaningful hand movements, and left IFG more sensitive to meaningful movements). Finally, in prior work in our lab [Skipper et al., 2007a], we used the same stimuli used in this study and a subset ( $N = 12$ ) of the participants from the current study, and found that IFG was an important component of a network linking gesture with speech. That investigation used structural equation modeling (SEM) to explore the influence of inferior frontal and premotor regions on temporal and inferior parietal regions when participants listened to stories that included meaningful gestures (*Gesture*), meaningless hand movements (*Self-Adaptor*), information from the face only (*No-Hand-Movement*), or no speech-related visual input (*No-Visual-Input*). When the hand movements were related to the

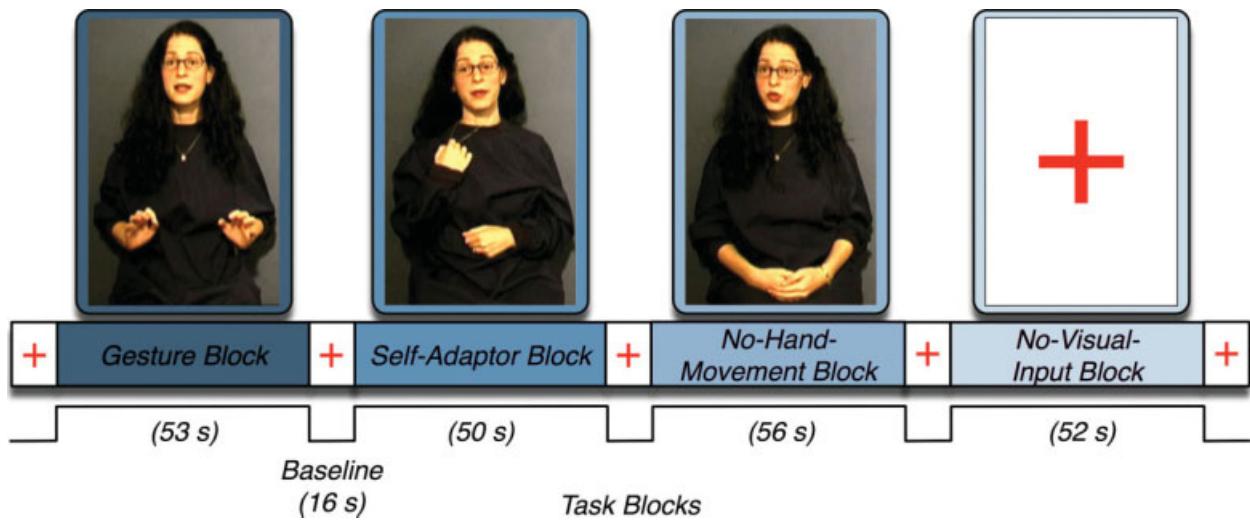
accompanying speech (*Gesture*), IFGTr and *pars opercularis* (IFGOp) of IFG exhibited *weaker* influence on other motor and language relevant cortical areas than when the hand movements were meaningless (*Self-Adaptor*) or when there were no accompanying hand movements (*No-Hand-Movement* or *No-Visual-Input*).

### The Present Study

Our prior work [Skipper et al., 2007a] focused on the role of IFG in processing gesture, but left some questions unanswered. It did not address differences in activity in specific brain regions, including IFG and STSp, and for purposes of SEM, data were collapsed across hemisphere, leaving open the selective roles of right and left hemisphere regions. Semantic information in language is known to be processed diffusely by regions in both the left and right hemispheres [Bookheimer, 2002; Ferstl et al., 2008; Jung-Beeman, 2005], but the contribution of both hemispheres to processing gesture has not been extensively investigated. In addition, we did not investigate in STSp differences in the response to meaningful and non-meaningful hand movements in the way that we do here. Holle et al. [2008] suggest that this region is sensitive to the semantic information from gesture, a finding that conflicts with prior investigations suggesting STSp is not primarily involved in processing semantic contributions from gesture [Willems et al., 2007].

This study thus examines in particular the role of bilateral STSp and IFG in integrating gesture and speech, and attempts to clarify the disparate neurobiological findings on gesture. In order to increase power, we recruited 12 additional participants, extending our prior sample [Skipper et al., 2007a]. To understand how these neural systems respond when presented with language in a relatively natural or “real-world” context, we used audiovisual stimuli containing manual gestures and movements of the face, lips, and mouth, a significant source of phonological information [Ross et al., 2007; Skipper et al., 2005, 2007b]. As in our previous work, participants listened to stories under four conditions: (1) with meaningful hand and arm gestures (i.e., *Gesture*), (2) with self-adaptive (grooming) hand movements that are not intentionally produced to communicate [Ekman and Friesen, 1969] and, in our case, bore no meaningful relation to speech (i.e., *Self-Adaptor*), (3) without manual gesture but with oral and facial gesture (i.e., *No-Hand-Movement*), and (4) without speech-related visual information (i.e., *No-Visual-Input*; see Fig. 1).

Using this experimental manipulation, we examined how the brain processes information about the actions of others, and integrates the semantic features of these biological movements to speech. To this end, we had two main goals. First, to assess the functional role of STSp, we contrasted conditions with hand movements (*Gesture* and *Self-Adaptor*) to those without hand movements (*No-Hand-Movement* and *No-Visual-Input*). Some works suggests STSp

**Figure 1.**

Time-course of the experiment for one of the two runs. Shown are still pictures from videos corresponding to each of the language conditions. The blank screen with a fixation cross was shown during the *No-Visual-Input* and the *Baseline* conditions.

responds to both meaningful and non-meaningful hand movements [Hubbard et al., 2009; Willems et al., 2007; Wilson et al., 2008], but other work [Holle et al., 2008] suggests that it is more sensitive to meaningful gestures. To assess the semantic contribution of gesture, our second contrast focused on conditions with and without meaningful hand movements (i.e., *Gesture* vs. *Self-Adaptor*). On the basis of our prior work [Skipper et al., 2007a], and on those of Willems et al. [2007], who found greater activity for incongruent gestures, we expected greater activity in IFG for non-meaningful hand movements than for meaningful gestures. Given the recent evidence suggesting that anterior and posterior IFG can be functionally dissociated [Gold et al., 2005; Snyder et al., 2007], we also expected anterior IFG to be most sensitive to this contrast. In addition, because narrative-level language has been shown to recruit bilateral IFG [Ferstl et al., 2008; Skipper et al., 2005; Wilson et al., 2008], and because Straube et al. [2008] reported right IFG was sensitive to gesture, we investigated the response in IFG in both hemispheres. The results of this study will thus help constrain neurobiological theories of gesture by distinguishing how STSp and IFG respond to hand movements that accompany speech, and which brain regions are particularly important for integrating semantic aspects of gesture with accompanying speech.

## MATERIALS AND METHODS

### Participants

Twenty-four adults (12 females,  $M$  age = 23.0 years,  $SD$  = 5.6 years) participated: 12 of these participants were

part of a prior investigation conducted in our laboratory, which used the same experimental manipulation [Skipper et al., 2007a]. All participants reported being right-handed, used their right hand for writing, had a positive Edinburgh handedness score [Oldfield, 1971], and had normal hearing and normal (or corrected normal) vision. All were native English speakers without early exposure to a second language. No participant had a history of neurological or psychiatric illness. Participants gave written informed consent and the Institutional Review Board of the Biological Sciences Division of The University of Chicago approved the study.

### Stimuli

Participants completed two runs of four experimental conditions, each lasting approximately 5 min. Audio was delivered at a sound pressure level of 85 dB-SPL through MRI-compatible headphones (Resonance Technologies, Northridge, CA), and video stimuli was viewed through a mirror attached to the head coil that allowed them to see a screen at the end of the scanning bed. The four conditions were separated by a baseline condition (*Baseline*) during which the participants viewed a fixation cross (16 s). The Baseline also occurred at the beginning and the end of the run. In each condition, participants listened to modified versions of Aesop's Fables, lasting 53 s on average ( $SD$  = 3 s).

The mode of presentation of the fable varied across the four conditions. In the first condition (*Gesture*), participants listened to a female telling a story while watching the speaker make natural co-speech gestures. Her gestures were primarily metaphoric or iconic [McNeill, 1992] each bearing a meaningful relationship to the semantic content

of the speech it accompanied. In the second condition (*Self-Adaptor*), participants watched the same speaker performing self-grooming hand movements, e.g., adjusting her hair or her shirt or scratching, that were unrelated to the content of the speech. In the third and fourth conditions, participants watched a speaker whose hands were resting in her lap (*No-Hand-Movement*), or saw only a blank screen with a fixation cross (*No-Visual-Input*; Fig. 1).

In all audiovisual conditions, the speaker was framed in the same way from waist to head with sufficient width to allow full perception of all hand movements. Before the experiment, participants were instructed to pay attention to the audio- and audiovisual stories but were given no other specific instructions. The videos in the gesture condition were coded using a previously established system [McNeill, 1992]. Eighty-five percent of the gestures presented to the participants were either iconic or metaphoric (movements that either captured a concrete physical characteristic of an object or an abstract notion). The remaining gestures were categorized as beat or deictic gestures, and none were categorized as codified emblems [Ekman and Friesen, 1969]. To control the nature of the actress's speech production and to match the number of hand movements in the *Gesture* and *Self-Adaptor* conditions, all stimuli were rehearsed. The actress also practiced the stimuli so that her prosody and lexical items were the same, occurring in similar temporal locations across the four sets of stimuli. To create the *No-Visual-Input* condition, we removed the video track from the *Gesture* condition.

One story from each condition was randomly presented in a run; each participant was exposed to two runs lasting approximately 5 min each. Participants thus heard a total of eight stories, two in each condition, and never heard the same story twice. Following each run, participants responded to true/false questions about each story using a button box (two per condition per run). Mean accuracy across participants was 85% (range 82–89%), with no significant difference in accuracy among conditions, suggesting that participants paid attention to the stories in all conditions.

### Data Collection

MRI scans were acquired on a 3-Tesla scanner with a standard quadrature head coil (General Electric, Milwaukee, WI). Volumetric T1-weighted scans (120 axial slices, 1.5 mm × 0.938 mm × 0.938 mm resolution) provided high-resolution anatomical images. For the functional scans, 30 slices (voxel size 5.00 mm × 3.75 mm × 3.75 mm) were acquired in the sagittal plane using spiral blood oxygen level dependent (BOLD) acquisition [TR/TE = 2000 ms/25 ms, FA = 77°; Noll et al., 1995]. The same acquisition was repeated continuously throughout each of the two functional runs. The first four BOLD scans of each run were discarded to avoid images acquired before the signal reached a steady state.

### Data Analysis

We used the Freesurfer (<http://surfer.nmr.mgh.harvard.edu>), Analysis of Functional Neuroimages/Surface Mapping with AFNI (AFNI/SUMA; <http://afni.nimh.nih.gov>), R statistical (<http://www.R-project.org>), and MySQL (<http://www.mysql.com/>) software packages for all analyses. For each participant, functional images were registered to the first non-discarded image of the first run and co-registered to the anatomical volumes, and multiple linear regression was performed individually for each voxel-based time series [Cox, 1996]. To model sustained activity across all time points within a block, regressors corresponding to the four different language conditions were convolved with a gamma function model of the hemodynamic response derived from Cohen [1997]. We included the results of this convolution as predictors along with the mean, linear and quadratic trends, and six motion parameters, obtained from the spatial alignment procedure. This analysis resulted in individual regression coefficients (beta weights) corresponding to each language condition, implicitly contrasted to the resting baseline, and *t* statistics measuring the reliability of the associated regression coefficients.

We conducted second level group analysis on a two-dimensional surface rendering of the brain [Dale et al., 1999; Fischl et al., 1999a], first analyzing whole-brain activation on the cortical surface and then probing activity in individual regions of interest (ROIs). We chose a surface-based approach because, in contrast to volume-based approaches, surface-based analyses result in data that more accurately reflect the cortical folding pattern of individual brains, reducing intersubject variability and increasing statistical power [Argall et al., 2006]. To initiate this surface approach, we used Freesurfer to (a) segment the white and grey matter of anatomical volumes, (b) inflate the cortical surfaces separately for each hemisphere [Dale et al., 1999; Fischl et al., 1999a], and (c) register each hemisphere to a common template of average curvature to optimally align the sulcal and gyral features across participants while minimizing individual metric distortion [Fischl et al., 1999b].

We converted (i.e., "normalized") the regression coefficients for each voxel to reflect percent signal change and interpolated these values from the three-dimensional volume domain to the two-dimensional surface representation of each individual participant's anatomical volume using SUMA. This program mapped the normalized coefficient and associated *t* statistics to a specific node on the surface representation of the individual's brain. ROI analyses were conducted on these individual surfaces. Image registration across the group required an additional step to standardize the number of nodes across individuals, and this was accomplished in SUMA using icosahedral tessellation and projection [Argall et al., 2006]. To decrease spatial noise for all analyses, the data were smoothed on the surface using a heat kernel of size 4 mm FWHM [Chung et al.,

2005]. Smoothing on the surface as opposed to the volume insures that white matter values are not included, and that functional data situated in anatomically distant locations on the cortical surface are not averaged across sulci [Argall et al., 2006; Desai et al., 2005]. We imported smoothed values of each surface node from each individual to tables in a MySQL relational database, with each table specifying a hemisphere of individual brains. The R statistical package was then used to query the database and analyze the information stored in these tables [for details see Hasson et al., 2008]. Finally, we used Freesurfer to create an average of the individual cortical surfaces on which to display the results of the whole-brain analysis.

### **Signal-to-noise ratio (SNR)**

We carried out a SNR analysis to determine if there were any cortical regions where, across participants, it is impossible to find experimental effects simply due to high noise levels [see Parrish et al., 2000 for rationale of using this method in fMRI studies]. We present the details of this analysis in the Supporting Information and below relate it to specific results.

### **Whole-brain analysis**

A mixed (fixed and random) effects Condition (4)  $\times$  Participant (24) analysis of variance (ANOVA) was conducted on a node-by-node basis using the normalized coefficients (i.e., signal change) from each individual's regression model as the dependent variable. Comparisons with the resting baseline were conducted for each condition, and five post-hoc comparisons among conditions were specified to explore our research questions. We controlled for statistical outliers on a node-by-node basis by removing those with signal change values greater than three-standard deviations from the mean signal of that in the transverse temporal gyrus, a region expected to include the primary auditory cortex, and thus to have reliable and robust activation in all conditions (outliers accounted for less than 1% of the data in each hemisphere).

To control for the family-wise error rate given multiple comparisons, we clustered the data using a nonparametric permutation method; a method that proceeds by resampling the data under the null hypothesis without replacement and that does not make any assumptions about the distribution of the parameter in question [see Hayasaka and Nichols, 2003; Nichols and Holmes, 2002 for implementation details]. Using this method, one can determine cluster sizes that would occur purely by chance, and in doing so determine a minimum cluster size to control for the family-wise error (FWE) rate (e.g., taking cluster sizes above the 95th percentile of the random distribution controls for the FWE at the  $P < 0.05$  level). For comparisons with baseline (*Gesture*, *Self-Adaptor*, *No-Hand-Movement*, and *No-Visual-Input* vs. resting baseline), the permutation method determined that a minimum cluster size of 262

surface nodes was sufficient to achieve a FWE rate of  $P < 0.05$  given a per-voxel threshold of  $P < 0.01$ , with clustering performed separately for positive and negative values. For comparisons assessing differences between conditions (*Gesture* vs. *No-Hand-Movement*, *Gesture* vs. *No-Visual-Input*, and *Gesture* vs. *Self-Adaptor*; *Self-Adaptor* vs. *No-Hand-Movement* and *Self-Adaptor* vs. *No-Visual-Input*), the permutation method determined a minimum cluster size of 1,821 surface nodes to maintain a FWE rate of  $P < 0.05$  given a per-voxel threshold of  $P < 0.05$ . The choice of these individual-voxel thresholds was motivated by the fact that our design, as shown in the SNR analysis (Supp. Info.), had more power to detect differences from absolute baseline (e.g., on the order of 1%) than to detect differences between conditions (even assuming a large 0.5% signal change difference).

The goal of the analysis was to characterize the regions most sensitive to the perception of hand movements (meaningful or not), and those most sensitive to hand movements that have a meaningful relation to speech. To address hand motion perception we examined the intersection ["conjunction"; Nichols et al., 2005] of brain activity during the *Gesture* and *Self-Adaptor* conditions. Intersection maps were created to separately assess activation specific to hand movements relative to *No-Hand-Movement* and *No-Visual-Input*, using thresholded direct contrasts across conditions: (a) [*Gesture* vs. *No-Hand-Movement*]  $\cap$  [*Self-Adaptor* vs. *No-Hand-Movement*] and (b) [*Gesture* vs. *No-Visual-Input*]  $\cap$  [*Self-Adaptor* vs. *No-Visual-Input*]. This analytic approach is equivalent to calculating the logical AND (intersection) among active nodes in each comparison, requiring that all comparisons within the conjunction be individually significant. To determine the question of which brain regions were sensitive to the semantic relationship between gesture and speech, we used the direct whole-brain contrast, *Gesture* vs. *Self-Adaptor*.

### **Region of interest analyses**

We tested the hypotheses outlined in the introduction—that co-speech gestures ought to affect the hemodynamic response in STSp, IFGOp, and IFGTr—by investigating these specific anatomically defined ROIs. These ROIs (detailed in the Results section and in Figs. 3 and 4) were defined on each individual surface representation using an automated parcellation procedure [Desikan et al., 2006; Fischl et al., 2004]. This procedure utilizes a probabilistic labeling algorithm incorporating the neuroanatomical conventions of Duvernoy [1991] and has a high accuracy approaching the accuracy of manual parcellation [Desikan et al., 2006; Fischl et al., 2002, 2004]. Although the default parcellation includes a demarcation of the three parts of IFG (IFGOp, IFGTr, and *pars orbitalis*; IFGOr), it does not demarcate the superior temporal cortex into anterior and posterior parts. We therefore modified this anatomical parcellation scheme to include a STSp, which was defined as all STS posterior to Heschl's gyrus. This definition was used by [Saygin et al., 2004] in their study of point-light biological motion perception.

Data analysis proceeded separately for each ROI (STSp, IFGOp, and IFGTr) using the percent signal change as the dependent variable. As a first step in the analysis, nodes within each ROI that contributed outlying normalized beta values (defined as greater than three  $SDs$  away from the mean of the ROI for that participant) were removed. We then checked for participant outliers and found that, in the *No-Visual-Input* condition of both right inferior frontal ROIs, one participant contributed a mean percent signal change value greater than 6  $SDs$  from the mean of all participants for that condition. Such an extreme data point has the potential to drastically affect the results; we therefore removed this participant from all ROI analyses (n.b., because contrasts were computed on a node-wise basis, outliers were identified and removed on a node-wise basis in the whole-brain analysis).

To determine if any observed differences provided by the statistical comparisons in the whole brain were due to differences in positive activation of the relevant areas, or instead, due to deactivation in a comparison condition, data were thresholded to include only the top 25% of positive percent signal change values [Mitsis et al., 2008]. As an additional step to facilitate comparison across ROIs, we scaled percent signal change by the number of surface nodes contributing data within that ROI. In lieu of omnibus ANOVAs, we assessed our questions by focusing on specific conditions of interest with a priori repeated measures contrasts [Rosenthal et al., 2000] conducted for each hemisphere separately (detailed below).

## RESULTS

We first briefly present the results of the SNR analysis as they indicated that certain regions, typically associated with susceptibility artifacts in fMRI studies, also demonstrated low SNR in the current experiment. For this reason, we do not discuss them further. Second, we present the general results on those regions that showed either above or below baseline BOLD signal during language comprehension, and compare them to the prior literature on auditory and audiovisual discourse comprehension. Activation (or deactivation) was defined in reference to the 16 s baseline periods of visual fixation (Fig. 2 and Supp. Table I). We then present two sets of results related to our specific hypotheses. We present the results for the comparison of conditions with hand movements vs. without hand movements (*Gesture* and *Self-Adaptor* vs. *No-Hand-Movement* and *No-Visual-Input*), specifically assessing bilateral STSp ROIs (see Fig. 3). We also present the results for the comparison of conditions with versus without meaningful hand movements (*Gesture* vs. *Self-Adaptor*) in the whole-brain contrast, as well as our assessment of bilateral STSp and inferior frontal (IFGOp and IFGTr) ROIs (Figs. 3 and 4).

### **Signal-to-Noise Ratio**

Simulations indicated that in the current design, the minimum SNR needed to detect a signal change of 0.5% was 54, and that needed to detect a signal change of 1% was 27 (see Supp. Info). We analyzed the mean SNR

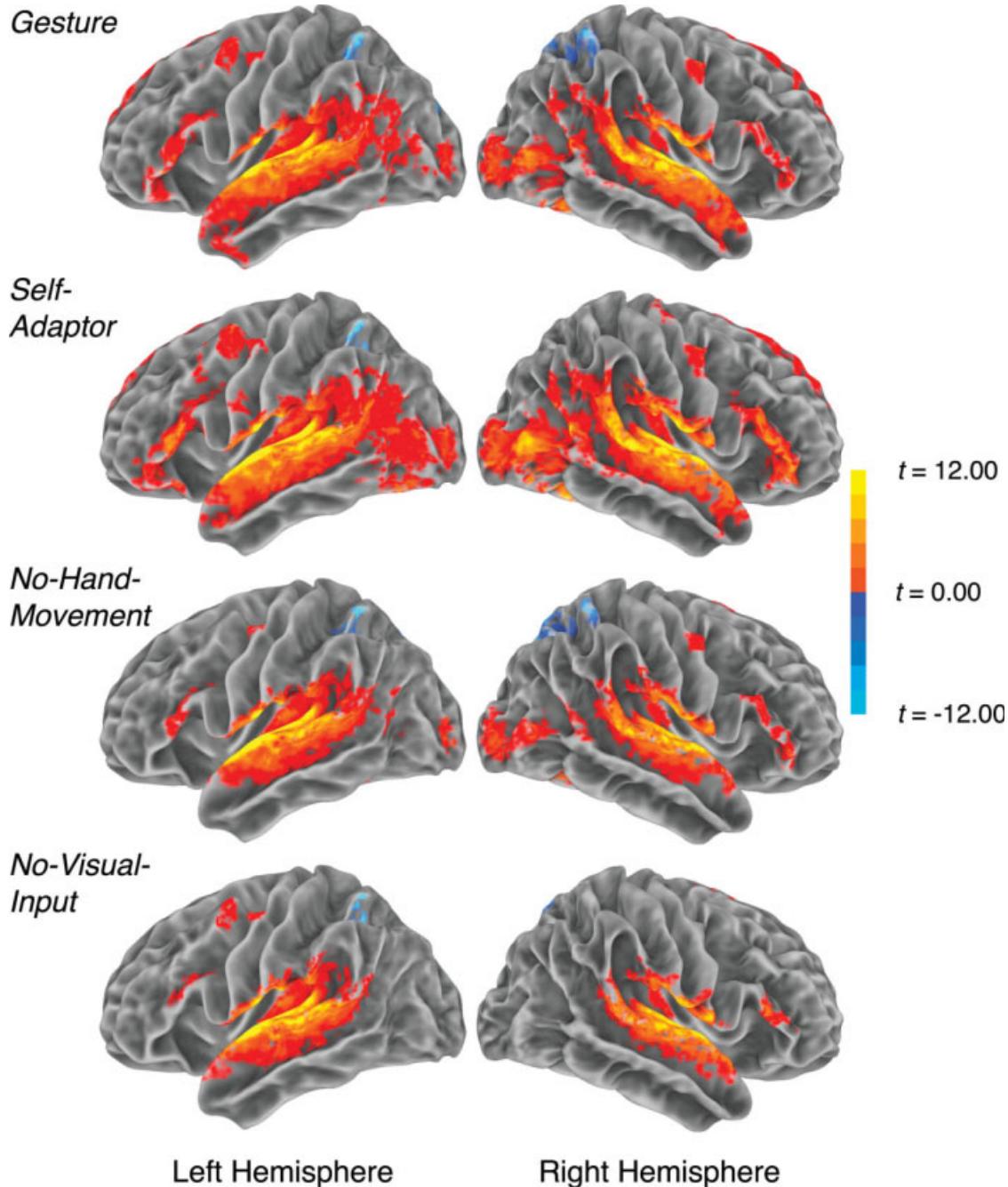
across participants from 18 ROIs in the inferior frontal, temporal, inferior parietal, and occipital cortex. In the regions we examined mean SNR ranged from a low 12.56 ( $SD = 7.03$ ) in right temporal pole (a region of high susceptibility artifact) to a high 93.15 ( $SD = 14.52$ ) in right calcarine sulcus. Importantly, in most regions SNR was sufficient to detect even small signal changes, and when collapsing over ROIs, SNR did not differ between hemispheres ( $M$  left = 59.25;  $M$  right = 59.12). For the regions we focused on (STSp, IFGOp, and IFGTr) there was adequate SNR to detect signal changes of at least 0.5% in both hemispheres (minimum SNR = 54.48 in left IFGTr). Although there were no significant left-right differences for STSp and IFGOp, SNR in right IFGTr was greater than in the left IFGTr, ( $70.71$  vs.  $54.48$ ,  $t[23] = -4.62$ ,  $p < 0.001$ ).

### **General Findings: Auditory and Audiovisual Discourse Comprehension Compared to Resting Baseline**

We first examined activation and deactivation (in which the voxel's time series is negatively correlated with the modeled idealized hemodynamic response function) in all four conditions relative to a resting baseline. In Supporting Table I, we present cluster size and maximum intensity of the clusters compared to baseline, as well as stereotaxic coordinates (in the volume space) and Brodmann Area of the center of mass for comparison with prior work. All four conditions contrasted with rest showed bilateral activation in frontal, inferior parietal, and temporal regions, findings that are comparable with prior studies and meta-analyses of speech- and discourse-level language comprehension [Ferstl et al., 2008; Indefrey and Cutler, 2004; Straube et al., 2008; Vigneau et al., 2006; Wilson et al., 2008]. Activation in occipital visual regions and in posterior fusiform was found in all three conditions containing visual information from the speaker (i.e., *Gesture*, *Self-Adaptor*, and *No-Hand-Movements*). Relative to baseline, the *Gesture* and *Self-Adaptor* conditions revealed activation across bilateral superior and medial frontal gyrus (extending across ~BA 6, 8, and 9; not visible in Fig. 2). Clusters negatively correlated with the modeled idealized hemodynamic response function (i.e., “deactivated” clusters) were found in posterior and medial structures bilaterally (with centers of mass in posterior cingulate, precuneus, and cuneus), and also in lateral superior parietal cortex and lingual gyrus. These deactivated clusters are consistent with those found in other discourse and language studies [Hasson et al., 2007; Wilson et al., 2008]. Thus, our findings of neural activity during discourse comprehension are consistent with those reported in the prior literature.

### **Biological Motion Perception and Co-Speech Gesture**

We explored regions that were sensitive to the biological motion aspects of co-speech gestures by comparing

**Figure 2.**

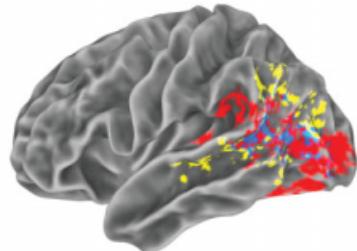
Whole-brain analysis results for each condition compared to *Baseline*. The individual per-voxel threshold was  $P < 0.01$  (corrected FWE  $P < 0.05$ ).

conditions with hand movements to those without hand movements. To investigate this question, we computed the intersection of the surface nodes that were statistically significant (corrected) above baseline corresponding to comparisons of interest across conditions. Figure 3A shows the results for (a) conditions with hand movements compared

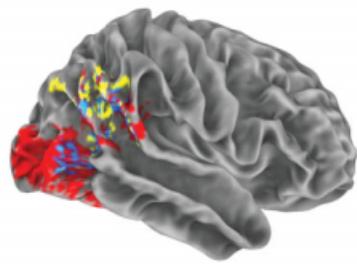
to the condition in which only visual information of the face and body was available (i.e., [Gesture vs. No-Hand-Movement]  $\cap$  [Self-Adaptor vs. No-Hand-Movement], shown in yellow); and (b) conditions with hand movements compared to the condition in which only visual information from a fixation cross was available (i.e., [Gesture vs.

## A Intersection Analysis

■ Hand-movements > Visual information from face  
 $[G > N-H-M] \cap [S-A > N-H-M]$

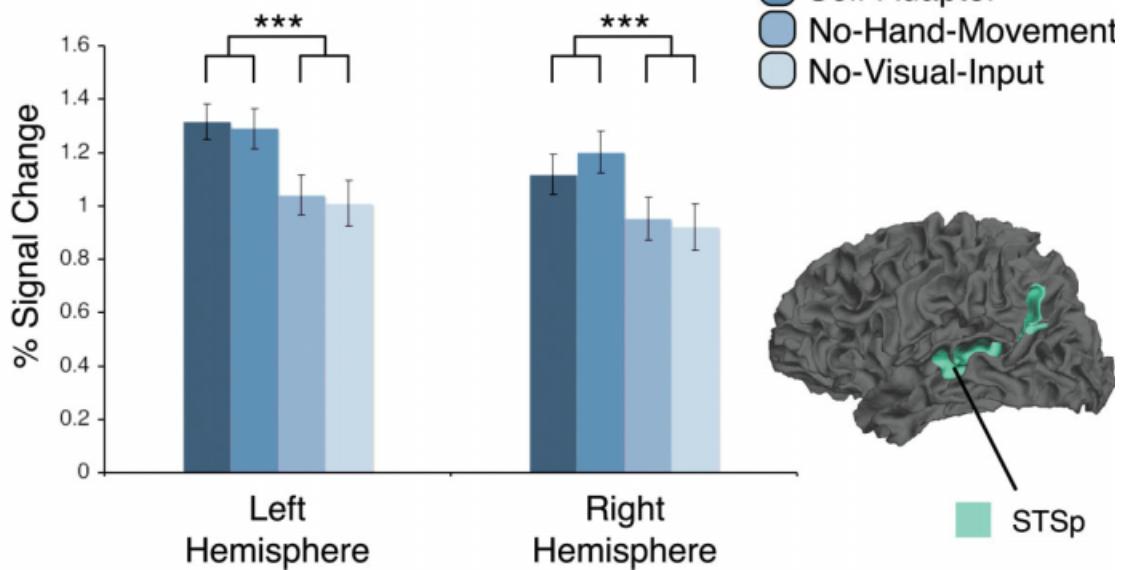


■ Hand-movements > No visual information  
 $[G > N-V-I] \cap [S-A > N-V-I]$



■ Hand-movements > Both

## B Region of Interest: STSp



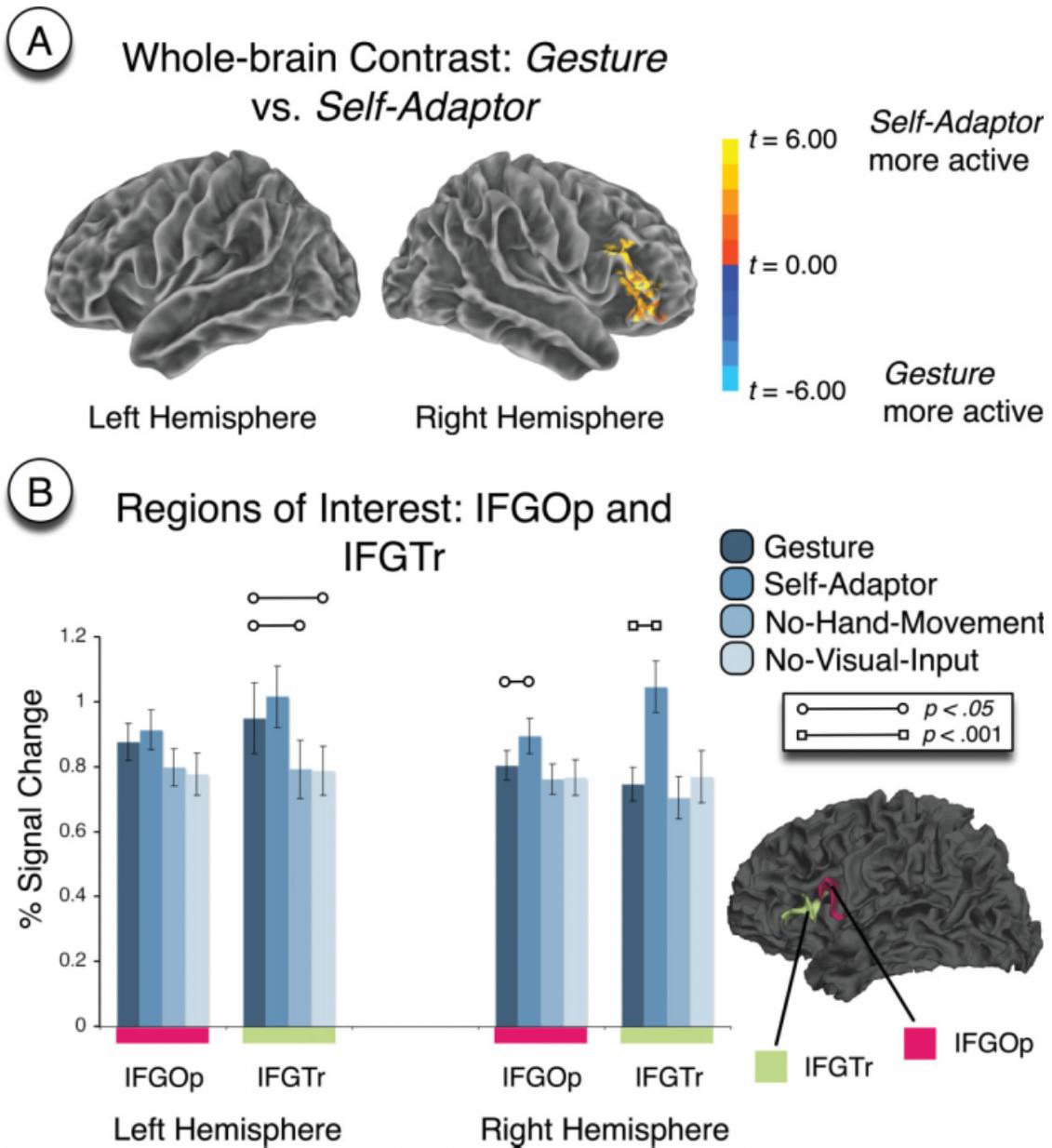
**Figure 3.**

(A) Results of the intersection analysis for conditions with hand movements relative to no hand movements. G, Gesture; S-A, Self-Adaptor; N-H-M, No-Hand-Movement; N-V-I, No-Visual-Input. (B) Results of the analysis for the posterior superior temporal sulcus (STSp) region of interest. \*\*\*  $P < 0.001$ .

No-Visual-Input]  $\cap$  [Self-Adaptor vs. No-Visual-Input], shown in red). Blue shows areas in which activity was present for both of these comparisons.

Compared to the No-Hand-Movement condition that included information from the face, both the Gesture and

Self-Adaptor conditions elicited more activity in bilateral STSp and STGp, IPL, middle occipital gyrus and sulcus, and left MTGp. Compared to the No-Visual-Input condition that contained only auditory input, both the Gesture and Self-Adaptor conditions elicited more activity bilaterally in

**Figure 4.**

(A) Results of the whole-brain comparison between *Self-Adaptor* and *Gesture*. Figure shows the left and right hemispheres (no difference was detected on the left). (B) Results of the analysis for inferior frontal regions of interest. Comparisons were conducted only between *Gesture* and all other conditions (differences between *Self-Adaptor* and *No-Hand-Movement* and *Self-Adaptor* and *No-Visual-Input*, and between *No-Hand-Movement* and *No-Vis-*

*ual-Input*, were not a focus and were not statistically assessed). Left *pars triangularis* (IFGTr) was sensitive to gestures relative to conditions without hand movements. Both right IFGTr and *pars opercularis* (IFGOp) were more active for *Self-Adaptor* than for *Gesture*, but notably, analyses of the interaction assessing the moderating influence of hemisphere was significant only for IFGTr.

early visual cortices (~BA 17, 18, 19), in bilateral posterior fusiform gyrus typically active when participants are viewing faces [Kanwisher et al., 1997; McCarthy et al., 1997], and in posterior cortical regions associated with processing

biological motion and in crossmodal integration of the auditory and visual modalities [i.e., IPL; STSp and STGp; and lateral occipitotemporal cortex; Calvert, 2001; Puce and Perrett, 2003]. Activation was also significantly greater

in left Heschl's gyrus, consistent with prior findings of crossmodal integration of information even in early auditory areas [Miller and D'Esposito, 2005].

We next probed activity in the STSp anatomical ROI, which was identified on each individual surface representation for each hemisphere (see Fig. 3B inset), with focused comparisons. We collapsed the data and assessed conditions with hand movements compared to those without hand movements ([*Gesture* and *Self-Adaptor*] vs. [*No-Hand-Movement* and *No-Visual-Input*]; as detailed in the section later, no differences between *Gesture* and *Self-Adaptor* were found in either hemisphere). Activity in STSp was significantly greater for conditions with hand movements in both hemispheres (left:  $t[22] = 6.56, p < 0.001$ ; right:  $t[22] = 5.42, p < 0.001$ ). Taken together, the results of both the intersection analysis and ROI analysis suggest that bilateral STSp, in conjunction with an extensive posterior cortical network, is centrally involved in processing the biological motion aspects of hand movements accompanying speech.

### Semantic Processing and Co-Speech Gesture

In this analysis we assessed, on the whole brain, which regions were sensitive to the semantic relationship between hand movements and accompanying speech by comparing the *Gesture* condition to the *Self-Adaptor* condition. We found only one significant cluster covering primarily the anterior subdivisions of right IFG (IFGOr and IFGTr; Talairach center of mass in the volume domain for cluster;  $x = 46$  mm,  $y = 30$  mm,  $z = 5$  mm, see Fig. 4A). In these regions the *Self-Adaptor* condition elicited greater activation than the *Gesture* condition.

We next conducted analyses for specific inferior frontal ROIs IFGTr and IFGOP, the results of which are shown in Figure 4B, with a representative surface of an individual participant presented in the inset. The a priori comparisons specified for these regions were *Gesture* vs. *Self-Adaptor*, *Gesture* vs. *No-Hand-Movement*, and *Gesture* vs. *No-Visual-Input*. Because prior research has found greater activity in IFG during the observation of hand movements and gestures [Kircher et al., 2009; Molnar-Szakacs et al., 2005; Willems et al., 2007], we used one-tailed significance tests for the *Gesture* vs. *No-Hand-Movement* and *Gesture* vs. *No-Visual-Input* comparisons with the expectation that *Gesture* would elicit greater activity in both comparisons. For reasons outlined in the Introduction, we also used one-tailed significance tests for the *Gesture* vs. *Self-Adaptor* comparison with the expectation that *Self-Adaptor* would elicit greater activity.

Several significant differences were found for IFGTr. In the left hemisphere (LH), *Gesture* was significantly greater than both *No-Hand-Movement* and *No-Visual-Input*,  $t(22) = 1.69, P < 0.05$  (one-tailed),  $d = 0.72$  and  $t(22) = 2.39, P < 0.05$  (one-tailed),  $d = 1.02$  respectively, but no strong difference was found for *Self-Adaptor* > *Gesture* ( $P = 0.30$

[one-tailed];  $d = 0.23$ ). In contrast, in the RH IFGTr, the comparison *Self-Adaptor* > *Gesture* was significant,  $t(22) = 4.16, P < 0.001, d = 1.77$ , which is also consistent with the whole-brain analysis. However, no significant differences were found for the other two contrasts, *Gesture* > *No-Hand-Movement* and *Gesture* > *No-Visual-Input* (largest  $t[22] = 0.63, P > 0.05$  [one-tailed],  $d = 0.27$ ). These findings point to the possibility that the response in these brain regions is moderated by hemisphere. We statistically tested this moderating influence of hemisphere by assessing the interaction of hemisphere by condition using the approach outlined in Jaccard [1998]. We found that the contrast of *Self-Adaptor* > *Gesture* was reliably stronger in RH than in LH, resulting in a reliable interaction term (i.e., right [*Self-Adaptor-Gesture*] > left [*Self-Adaptor-Gesture*],  $t[22] = 1.90, P < 0.05, d = 0.81$ ). The contrast of *Gesture* > *No-Visual-Input* was reliably stronger in LH than in RH (left [*Gesture-No-Visual-Input*] > right [*Gesture-No-Visual-Input*],  $t[22] = 2.30, P < 0.05, d = 0.98$ ), but the contrast *Gesture* > *No-Hand-Movement* was not (left [*Gesture-No-Hand-Movement*] > right [*Gesture-No-Hand-Movement*],  $t[22] = 1.29, P > 0.05, d = 0.55$ ). Thus, interpreted in the context of the interactions, the right, but not left IFGTr, was sensitive to the semantic manipulation. The left, but not right IFGTr, was sensitive to meaningful gestures relative to auditory-only discourse.

For IFGOP, the only significant difference for any contrast in either hemisphere was for the *Self-Adaptor* > *Gesture* contrast in the RH,  $t(22) = 1.95, P < 0.05$  (one-tailed),  $d = 0.83$ , but the hemisphere by condition interaction was found to be nonsignificant (right [*Self-Adaptor-Gesture*] > left [*Self-Adaptor-Gesture*],  $P = 0.21, d = 0.36$ ), and thus we cannot make any strong claims about hemispheric differences in IFGOP. Later and in the Discussion, we interpret all results for IFG taking into account the significant interaction effects.

Because Holle et al. [2008] reported greater activity for iconic gestures relative to grooming movements in STSp, we also assessed the *Gesture* versus *Self-Adaptor* contrast in this anatomical ROI. A significant difference favoring *Gesture* would have been consistent with the findings reported by Holle et al. However, we found no significant difference in either hemisphere (largest  $t[22] = 1.56, P > 0.05$  (two-tailed),  $d = 0.66$  favoring *Self-Adaptor*; note we also found no difference for the *Gesture* vs. *Self-Adaptor* comparison in the whole-brain analysis). Thus, STSp was not found to be sensitive to the semantic relation between the hand movement and the accompanying speech. This result fails to replicate Holle et al. [2008] who found significantly greater activity in this region for iconic gestures than for grooming movements.

Importantly, the SNR analysis is relevant to our confidence in the results reported earlier. This is because SNR not only speaks to our ability to detect activation relative to resting baseline, but also to the ability to obtain an accurate and reliable beta estimate (i.e., percent signal change estimate) for each participant and for the participant group

as a whole. Simply put, the assessment of signal change in regions with high SNR is more accurate (reliable), and we can be confident in the findings in those regions. Thus, because SNR in STSp and IFG were quite sufficient to detect even small signal changes, we can be confident in our null finding in STSp, and in our findings of significant differences in left and right IFG. However, the SNR analysis also suggests caution in interpreting the hemispheric differences in IFGTr, where higher SNR on the right might have led to more stable estimates of percent signal change across participants, and might explain why the semantic manipulation was found only in right IFG, and was only approaching significance in the LH. Indeed, when we looked at the data from the 12 participants with the highest SNR values (the top half of the sample), the difference between *Gesture* and *Self-Adaptor* was more robust if not significant ( $P < 0.06$ ). This suggests future work on the role of IFG in gesture processing should strongly consider methods to increase SNR to detect subtle responses in this region to semantic information from gestures.

To summarize, interpreted in the context of the SNR and interaction analyses, we found that while left IFGTr responded more strongly to gestures during audiovisual discourse relative to auditory-only discourse, it was not more active for meaningless hand movements than for meaningful hand movements. Right IFGTr and (to a lesser degree) right IFGOp showed the opposite response. We found more activity for meaningless hand movements than for meaningful hand movements, but not for gestures during audiovisual discourse compared to auditory-only discourse. Finally, STSp was not sensitive to the semantic manipulation-no differences between *Gesture* and *Self-Adaptor* were found in this region.

## DISCUSSION

The goal of this study was to explore the integration of gesture and speech, focusing in particular on STSp's role in processing hand movements that accompany speech, and IFGs' role in processing the meaningfulness of those movements. We found that when hand movements accompany speech (i.e., *Gesture* and *Self-Adaptor* conditions) activity increased in regions typically associated with auditory language comprehension, which were also active for auditory-only spoken language in the current study. These included left IFG (indicated in the ROI analysis), left primary auditory cortex, and bilateral STGp, MTGp, STSp, and IPL. Other regions, which were independently identified as being sensitive to mouth movements during audiovisual language, were also sensitive to hand movements. These included bilateral STSp, a region associated with the perception of biological motion, and with the integration of auditory and visual information. Importantly, these posterior cortical regions were not sensitive to the semantic relation of gesture and speech, i.e., they did not distinguish those hand movements that were semantically

related to the accompanying speech (*Gesture*) from those that were not (*Self-Adaptor*). Instead, right IFG was sensitive to the semantic relation between gesture and speech, becoming more active when hand movements were meaningless. The left anterior IFG did not differentiate between meaningful and meaningless hand movements, although it did respond more to speech with gestures than to speech without gestures. These findings suggest that bilateral IFG, in concert with bilateral posterior temporal and inferior parietal brain regions, comprise a distributed cortical network for integrating gestures with the speech it accompanies.

### Convergence of Auditory (Discourse) and Visual (Gesture) Information in Posterior Superior Temporal Sulcus

We found that bilateral STSp was more active when discourse was accompanied by hand movements compared to when it was not, but it did not differentiate hand movements that were meaningful (*Gesture*) versus those that were not (*Self-Adaptor*). *Prima facie*, this would seem to contrast with Holle et al. [2008], who reported greater STSp activity for meaningful iconic gestures than for non-meaningful grooming movements, a result suggesting sensitivity to the semantic information provided by gesture. Note, however, that the meaningful gestures in Holle et al.'s study were necessary to restrict the meaning of a homonym in the sentence. It is possible that participants were paying particular attention to gestures compatible with one of the homonym's meanings, a process that might increase activity in areas specialized for motion processing, but would likely not take place when listening to the type of naturalistic stimuli that we used in our study.

In our study, gesture accompanied sentences in a narrative discourse, which changes the way gesture is processed [McNeill et al., 1994, 2001]. To use Holle et al.'s terminology, in addition to local gesture-speech integration, our stimuli required global discourse-level integration, or integration of semantic constituents (words and gestures) with each other over time within discourse. Indeed, rather than occurring with a specific target homonym,<sup>1</sup> in our study gestures often preceded or followed the speech they were thematically related to, were not always tied to specific words, and often added information that was not given linguistically [e.g., path or speed of motion; Chui, 2005; Goldin-Meadow and Singer, 2003; Morrel-Samuels and Krauss, 1992]. It is possible that STSp responds differently

<sup>1</sup>Holle et al. audiovisual stimuli was edited to synchronize the auditory stream with the video stream such that the stroke of the gesture preceded or ended at the phonological peak of the syllable. The stimuli used in this study were not edited to affect the synchronization of gestures with speech.

to gesture when it is integrated with speech at the global discourse level as opposed to the local word level [c.f., Holle et al., 2008]. This interpretation finds some support from Willems et al. [2007] that left STSp was more active when the verb mismatched the sentence context, but not when the accompanying gesture mismatched the global level sentence interpretation, suggesting that meaningful gesture is processed in STSp differently in different linguistic contexts.

Our results support a role for STSp in processing biological motion (i.e., hand movements) whether or not those movements are meaningful. As such, the findings are consistent with prior work on biological motion processing in STSp [Beauchamp et al., 2002, 2003; Grossman et al., 2000; Thompson et al., 2005]. STSp has also been implicated, both anatomically and functionally, in the cross-modal integration of auditory and visual information [Calvert, 2001; Seltzer and Pandya, 1978, 1994], particularly with respect to speech [Calvert and Campbell, 2003; Calvert et al., 2000; Hocking and Price, 2008; Sekiyama et al., 2003; Skipper et al., 2005; Wright et al., 2003] and actions [Barraclough et al., 2005]. However, because in our study we did not have a visual-only comparison condition, the possibility remains open that STSp activity simply reflects processing of biological motion and does not play a special role in connecting gesture with speech. These open questions whether STSp integrates biological motion with speech, and whether STSp responds differently to gesture as a function of its linguistic context provide interesting avenues for future study.

### **Processing of Semantic Information From Co-Speech Gesture in the Inferior Frontal Gyrus**

We found that IFG is bilaterally involved in processing co-speech gestures with speech, but this finding can be refined by further anatomical specification. Our findings for left IFG indicate that this region demonstrates increased sensitivity to gestures in its anterior portion compared to the posterior portion. In particular, IFGTr responded more strongly when speech was accompanied by hand movements than when it was not. Our findings are consistent with recent work showing IFGTr is active during observation, but not imitation, of hand actions [Molnar-Szakacs et al., 2005], and with work suggesting IFG is important for processing multisensory auditory and visual stimuli [see Romanski, 2007 for review]. In addition, the growing gesture literature has suggested that the anterior portions of IFG are particularly important for integrating semantic information from gesture with speech [Skipper et al., 2007a; Straube et al., 2008; Willems et al., 2007], even when the gestures are congruent with the linguistic context [Kircher et al., 2009; Willems et al., 2007]. We also found evidence for a moderating influence of hemisphere. The contrast *Gesture > No-Hand-Movement* was found to be reliable in left IFG, but only the contrast of *Gesture > No-Visual-Input* was reli-

ably stronger in LH than in RH, providing some indication that left IFG was sensitive to co-speech gestures. However, left IFG was not sensitive to the semantic manipulation; right anterior IFG was most sensitive to the contrast of *Self-Adaptor > Gesture*.

Our results for IFG are important because they suggest a role for the right IFG in processing meaning from hand movements that accompany speech. Specifically, right IFG might play a role in the online revision of semantic interpretation because the information from self-adaptive grooming movements is difficult to integrate with speech into a coherent message. Indeed, if we consider that the goal of the listener is to understand the message of the speaker, every hand movement has the potential to contribute to the speaker's message for the listener, determining the relevance of the hand movement to the linguistic context is an online process. Evidence has shown that gestures are processed alongside speech in a somewhat automatic fashion [Holle and Gunter, 2007; Kelly et al., 2004; McNeill et al., 1994; Özyürek et al., 2007]. In addition, self-adaptors and incongruent gestures (but not congruent gestures) elicit increased N400 ERP components [Holle and Gunter 2007; Experiment 3], suggesting even self-adaptive hand movements influence semantic processing of speech. Considered in this context, it is possible that this difficulty of integration leads to greater reliance on language during this condition because grooming movements are meaningless in relation to the story. Alternatively, because grooming movements are difficult to integrate seamlessly into the auditory discourse, these movements might need to be ignored, an interpretation supported by studies suggesting one of the primary functions of right IFG is inhibition [see Aron et al., 2004 for a review].

The account we develop here calls for further evaluation, but it is consistent with prior findings in the literature. Although left IFG has been implicated in controlled retrieval [Gold and Buckner, 2002; Poldrack et al., 1999; Wagner et al., 2001] or selection of competing semantic meanings [Fletcher et al., 2000; Moss et al., 2005; Thompson-Schill et al., 1997], right IFG has only rarely been implicated specifically in semantic processes. This might, however, reflect a preoccupation with left hemisphere function for language.<sup>2</sup> Several studies of semantic selection in which left IFG was the focus also report activation in right IFG [Badre et al., 2005; Fletcher et al., 2000; Thompson-Schill et al., 1997], and right IFG is consistently implicated in the comprehension of figurative language [Lauro et al., 2008; Zempleni et al., 2007], linking of distant semantic relations [Rodd et al., 2005; Zempleni et al., 2007], and during semantic revision [Stowe et al., 2005]. With respect to audiovisual stimuli, two recent studies suggest that right IFG is sensitive to semantic conflict in

---

<sup>2</sup>It is notable that, while in their study of iconic co-speech gesture Willems et al. investigated bilateral inferior parietal and premotor ROIs, they only investigated a left inferior frontal ROI.

the two modalities. Hein et al. [2007] found that right IFG was more active during presentation of semantically incongruent audiovisual stimuli (e.g., a picture of a dog presented with a "meow" sound), but not during integration of semantically congruent stimuli (e.g., a picture of a dog presented with a "woof-woof" sound). Investigating gestures accompanying sentences, Straube et al. [2008] found that activation in bilateral IFG was correlated with correct recall of both previously presented metaphoric gestures, and meaningless hand movements, but that right anterior IFG was most strongly implicated in recall of meaningless hand movements. Finally, in our prior study of network-level connectivity [Skipper et al., 2007a], we did not focus on hemispheric differences, but we did find that IFG had a weaker impact on motor and language relevant cortices when speech was understood in the context of meaningful co-speech gestures as opposed to non-meaningful grooming movements. We interpreted this as reflecting the fact that gestures, contributing semantic information relevant to the message of the speaker, actually reduce ambiguity of the message and thus selection and retrieval demands (cf. Kircher et al., 2009]. For meaningful compared to meaningless gestures, the finding of both reduced connectivity with IFG in our prior study, and reduced activity in right IFG in this study, provides converging evidence to suggest IFG is sensitive to the relationship between accompanying hand movements and speech.

### **Observation–Execution Matching During Audiovisual Language Comprehension With Gesture**

A number of studies have identified brain regions, including ventral premotor cortex, IFG, and IPL, whose functional properties resemble those of macaque "mirror neurons," that fire during both execution of one's own actions and the observation of the same actions of others [for reviews see Iacoboni, 2005; Rizzolatti and Craighero, 2004; Rizzolatti et al., 2001]. It has been proposed that when perception is accompanied by gesture, this "observation-execution matching" process plays a role in disambiguating semantic aspects of speech by simulating motor programs involved in gesture production, i.e., the listener brings to bear upon observable gestures their own knowledge of the meaning of these gestures in part because cortical networks involved in producing them are active when they are perceived [Gentilucci et al., 2006; Holle et al., 2008; Iacoboni, 2005; Nishitani et al., 2005; Skipper et al., 2006; Willems and Hagoort, 2007; Willems et al., 2007].

Although our study did not include a gesture production condition, its results are relevant to the discussion of observation–execution matching and gesture. First, relative to baseline, bilateral premotor cortex was active in all conditions, which is comparable to that reported in other investigations of audiovisual speech perception and gesture [Calvert and Campbell, 2003; Holle et al., 2008;

Pulvermüller et al., 2006; Skipper et al., 2007b; Wilson and Iacoboni, 2006; Wilson et al., 2004]. Second, greater activation for hand movements in STSp, IPL, and IFG is consistent with studies showing STSp responds to congruency between an observed action and that same produced action [Iacoboni et al., 2001], and with studies showing IPL and IFG are active during observation of hand movements and meaningful gestures [Lotze et al., 2006; Molnar-Szakacs et al., 2005; Peigneux et al., 2004; Wheaton et al., 2004; Willems et al., 2007]. Although we cannot make strong claims about observation–execution matching and gesture, our findings does not support a straightforward mirror neuron account. Instead, they suggest that a more complete understanding of how the brain uses co-speech gestures to aid in understanding speech requires attention to superior temporal and inferior parietal regions in addition to the frontal (motor) regions often referenced in studies of observation–execution matching [see Skipper et al., 2006 for a similar view].

### **CONCLUSION**

Co-speech gestures serve an important role in language comprehension by providing additional semantic information that listeners can use to disambiguate the speaker's message. In this work, we have shown that a number of bilaterally distributed brain regions are sensitive to the additional information gestures contribute to the communication process. In particular, bilateral STSp was sensitive to hand movements but not to their semantic message. Bilateral IFG was also sensitive to gestures, but only right IFG distinguished between meaningful and non-meaningful hand movements. These findings show that perceiving hand movements during speech modulates the distributed pattern of neural activation involved in both biological motion perception and discourse comprehension, suggesting listeners attempt to find meaning, not only in the words speakers produce, but also in the hand movements that accompany speech.

### **ACKNOWLEDGMENTS**

We thank Michael Andric, Bernadette Brogan, Robert Fowler, Charlie Gaylord, Peter Huttenlocher, Susan Levine, Nameeta Lobo, Robert Lyons, Arika Okrent, Anjali Raja, Ana Solodkin, Linda Whealton-Suriyakam, and Lauren Wineburg.

### **REFERENCES**

- Argall BD, Saad ZS, Beauchamp MS (2006): Simplified intersubject averaging on the cortical surface using SUMA. *Hum Brain Mapp* 27:14–27.
- Aron AR, Robbins TW, Poldrack RA (2004): Inhibition and the right inferior frontal cortex. *Trends Cogn Sci* 8:170–177.

- Badre D, Poldrack RA, Pare-Blaauw EJ, Insler RZ, Wagner AD (2005): Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47:907–918.
- Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI (2005): Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cognit Neurosci* 17:377–391.
- Beattie G, Shovelton H (1999): Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *J Lang Soc Psychol* 18:438–462.
- Beauchamp MS, Lee KE, Haxby JV, Martin A (2002): Parallel visual motion processing streams for manipulable objects and human movements. *Neuron* 34:149–159.
- Beauchamp MS, Lee KE, Haxby JV, Martin A (2003): fMRI responses to video and point-light displays of moving humans and manipulable objects. *J Cognit Neurosci* 15:991–1001.
- Bonda E, Petrides M, Ostry D, Evans A (1996): Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J Neurosci* 16:3737–3744.
- Bookheimer S (2002): Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annu Rev Neurosci* 25:151–188.
- Calvert GA (2001): Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123.
- Calvert GA, Campbell R (2003): Reading speech from still and moving faces: The neural substrates of visible speech. *J Cognit Neurosci* 15:57–70.
- 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.
- Chui K (2005): Temporal patterning of speech and iconic gestures in conversational discourse. *J Pragmatics* 37:871–887.
- Chung MK, Robbins SM, Dalton KM, Davidson RJ, Alexander AL, Evans AC (2005): Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage* 25:1256–1265.
- Cohen MS (1997): Parametric analysis of fMRI data using linear systems methods. *NeuroImage* 6:93–103.
- Cohen RL, Otterbein N (1992): The mnemonic effect of speech gestures: Pantomimic and non-pantomimic gestures compared. *Eur J Cognit Psychol* 4:113–139.
- Cox RW (1996): AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Dale AM, Fischl B, Sereno MI (1999): Cortical surface-based analysis. I. Segmentation and surface reconstruction. *NeuroImage* 9:179–194.
- Desai R, Liebenthal E, Possing ET, Waldron E, Binder JR (2005): Volumetric vs. surface-based alignment for localization of auditory cortex activation. *NeuroImage* 26:1019–1029.
- Desikan RS, Segonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006): An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* 31:968–980.
- Duvernoy HM (1991): The Human Brain: Structure, Three-Dimensional Sectional Anatomy and MRI. New York: Springer-Verlag.
- Ekman P, Friesen WV (1969): The repertoire of nonverbal communication: Categories, origins, usage, and coding. *Semiotica* 1: 49–98.
- Ferstl EC, Neumann J, Bogler C, Yves von Cramon D (2008): The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Hum Brain Mapp* 29:581–593.
- Feyereisen P (2006): Further investigation on the mnemonic effect of gestures: Their meaning matters. *Eur J Cognit Psychol* 18:185–205.
- Fischl B, Sereno MI, Dale AM (1999a): Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *NeuroImage* 9:195–207.
- Fischl B, Sereno MI, Tootell RBH, Dale AM (1999b): High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8:272–284.
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale AM (2002): Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* 33:341–355.
- Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM (2004): Automatically parcellating the human cerebral cortex. *Cereb Cortex* 14:11–22.
- Fletcher PC, Shallice T, Dolan RJ (2000): Sculpting the response space—An account of left prefrontal activation at encoding. *NeuroImage* 12:404–417.
- Friederici AD, Hahne A, von Cramon DY (1998): First-pass versus second-pass parsing processes in a Wernicke's and a Broca's Aphasic: Electrophysiological evidence for a double dissociation. *Brain Lang* 62:311–341.
- Friederici AD, von Cramon DY, Kotz SA (1999): Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain* 122:1033–1047.
- Gentilucci M, Bernardis P, Crisi G, Dalla Volta R (2006): Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. *J Cognit Neurosci* 18:1059–1074.
- Gold BT, Buckner RL (2002): Common prefrontal regions coactivate with dissociable posterior regions during controlled semantic and phonological tasks. *Neuron* 35:803–812.
- Gold BT, Balota DA, Kirchhoff BA, Buckner RL (2005): Common and dissociable activation patterns associated with controlled semantic and phonological processing: Evidence from fMRI adaptation. *Cereb Cortex* 15:1438–1450.
- Goldin-Meadow S (2006): Talking and thinking with our hands. *Curr Dir Psychol Sci* 15:34–39.
- Goldin-Meadow S, Singer MA (2003): From children's hands to adults' ears: Gesture's role in the learning process. *Dev Psychol* 39:509–520.
- Goldin-Meadow S, Kim S, Singer M (1999): What the teacher's hands tell the student's mind about math. *J Educ Psychol* 91:720–730.
- Graham JA, Heywood S (1975): The effects of elimination of hand gestures and of verbal codability on speech performance. *Eur J Soc Psychol* 5:189–195.
- Grezes J, Fonlupt P, Bertenthal B, Delon-Martin C, Segebarth C, Decety J (2001): Does perception of biological motion rely on specific brain regions? *NeuroImage* 13:775–785.
- Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R (2000): Brain areas involved in perception of biological motion. *J Cognit Neurosci* 12:711–720.
- Hagoort P, Brown CM, Swaab TY (1996): Lexical-semantic event-related potential effects in patients with left hemisphere lesions and aphasia, and patients with right hemisphere lesions without aphasia. *Brain* 119:627–649.

- Hagoort P, Hald L, Bastiaansen M, Petersson KM (2004): Integration of word meaning and world knowledge in language comprehension. *Science* 304:438–441.
- Halgren E, Dhond RP, Christensen N, Van Petten C, Marinkovic K, Lewine JD, Dale AM (2002): N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage* 17:1101–1116.
- Hasson U, Nusbaum HC, Small SL (2007): Brain networks subserving the extraction of sentence information and its encoding to memory. *Cereb Cortex* 17:2899–2913.
- Hasson U, Skipper JI, Wilde MJ, Nusbaum HC, Small SL (2008): Improving the analysis, storage and sharing of neuroimaging data using relational databases and distributed computing. *NeuroImage* 39:693–706.
- Hayasaka S, Nichols TE (2003): Validating cluster size inference: random field and permutation methods. *NeuroImage* 20:2343–2356.
- 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.
- Helenius P, Salmelin R, Service E, Connolly JF, Leinonen S, Lyytinen H (2002): Cortical activation during spoken-word segmentation in non-reading impaired and dyslexic adults. *J Neurosci* 22:2936–2944.
- 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 Cognit Neurosci* 19:1175–1192.
- Holle H, Gunter TC, Rüschemeyer SA, Hennenlotter A, Iacoboni M (2008): Neural correlates of the processing of co-speech gestures. *NeuroImage* 39:2010–2024.
- Hubbard AL, Wilson SM, Callan DE, Dapretto M (2009): Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Hum Brain Mapp* 30:1028–1037.
- Iacoboni M (2005): Neural mechanisms of imitation. *Curr Opin Neurobiol* 15:632–637.
- Iacoboni M, Koski LM, Brass M, Bekkering H, Woods RP, Dubeau MC, Mazziotta JC, Rizzolatti G (2001): Reafferent copies of imitated actions in the right superior temporal cortex. *Proc Natl Acad Sci USA* 98:13995–13999.
- Indefrey P, Cutler A (2004): Prelexical and lexical processing in listening. In: Gazzaniga MS, editor. *The Cognitive Neurosciences*, 3rd ed. Cambridge: MIT Press. pp 759–774.
- Jaccard J (1998): Interaction Effects in Factorial Analysis of Variance. Thousand Oaks, CA: Sage.
- Jung-Beeman M (2005): Bilateral brain processes for comprehending natural language. *Trends Cognit Sci* 9:712–718.
- Kanwisher N, McDermott J, Chun MM (1997): The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.
- Kelly SD, Barr DJ, Church RB, Lynch K (1999): Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *J Mem Lang* 40:577–592.
- Kelly SD, Kravitz C, Hopkins M (2004): Neural correlates of bimodal speech and gesture comprehension. *Brain Lang* 89:253–260.
- Kelly SD, Ward S, Creigh P, Bartolotti J (2007): An intentional stance modulates the integration of gesture and speech during comprehension. *Brain Lang* 101:222–233.
- Kendon A (1988): How gestures can become like words. In: Poyatos F, editor. *Cross-Cultural Perspectives in Nonverbal Communication*. Toronto: Hogrefe & Huber Publishers. pp 131–141.
- Kendon A (1994): Do gestures communicate? A review. *Res Lang Soc Interac* 27:175–200.
- 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.
- Krauss RM, Dushay RA, Chen Y, Rauscher F (1995): The communicative value of conversational hand gestures. *J Exp Soc Psychol* 31:533–552.
- Kutas M, Federmeier KD (2000): Electrophysiology reveals semantic memory use in language comprehension. *Trends Cognit Sci* 4:463–470.
- Kutas M, Hillyard SA (1980): Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science* 207:203–205.
- Kwon H, Kuriki S, Kim JM, Lee YH, Kim K, Nam K (2005): MEG study on neural activities associated with syntactic and semantic violations in spoken Korean sentences. *Neurosci Res* 51:349–357.
- Lauro LJR, Tettamanti M, Cappa SF, Papagno C (2008): Idiom comprehension: A prefrontal task? *Cereb Cortex* 18:162–170.
- Lotze M, Heymans U, Birbaumer N, Veit R, Erb M, Flor H, Halsband U (2006): Differential cerebral activation during observation of expressive gestures and motor acts. *Neuropsychologia* 44:1787–1795.
- McCarthy G, Puce A, Gore JC, Allison T (1997): Face-specific processing in the human fusiform gyrus. *J Cognit Neurosci* 9:605–610.
- McNeill D (1992): *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press. 416 p.
- McNeill D (2005): *Gesture and Thought*. Chicago: University of Chicago Press. 318 p.
- McNeill D, Cassell J, McCullough K-E (1994): Communicative effects of speech-mismatched gestures. *Res Lang Soc Interac* 27:223–237.
- McNeil NM, Alibali MW, Evans JL (2000): The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *J Nonverbal Behav* 24:131–150.
- McNeill D, Quæk F, McCullough K-E, Duncan S, Furuyama N, Bryll R, Ma X-F, Ansari R (2001): Catchments, prosody and discourse. *Gesture* 1:9–33.
- Miller LM, D'Esposito M (2005): Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J Neurosci* 25:5884–5893.
- Mitsis GD, Iannetti GD, Smart TS, Tracey I, Wise RG (2008): Regions of interest analysis in pharmacological fMRI: How do the definition criteria influence the inferred result? *NeuroImage* 40:121–132.
- Molnar-Szakacs I, Iacoboni M, Koski L, Mazziotta JC (2005): Functional segregation within pars opercularis of the inferior frontal gyrus: Evidence from fMRI studies of imitation and action observation. *Cereb Cortex* 15:986–994.
- Morrel-Samuels P, Krauss RM (1992): Word familiarity predicts the temporal asynchrony of hand gestures and speech. *J Exp Psychol Learn* 18:615–622.
- Moss HE, Abdallah S, Fletcher P, Bright P, Pilgrim L, Acres K, Tyler LK (2005): Selecting among competing alternatives: Selection and retrieval in the left inferior frontal gyrus. *Cereb Cortex* 15:1723–1735.
- Nichols TE, Holmes AP (2002): Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Hum Brain Mapp* 15:1–25.

- Nichols T, Brett M, Andersson J, Wager T, Poline J-B (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.
- Noll DC, Cohen JD, Meyer CH, Schneider W (1995): Spiral K-space MR imaging of cortical activation. *J Magn Reson Imaging* 5:49–56.
- Oldfield RC (1971): The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia* 9:97–113.
- Olson IR, Gatenby JC, Gore JC (2002): A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cognit Brain Res* 14:129–138.
- Oram MW, Perrett DI (1994): Responses of anterior superior temporal polysensory (STPa) neurons to "biological motion" stimuli. *J Cognit Neurosci* 6:99–116.
- Oram MW, Perrett DI (1996): Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the Macaque monkey. *J Neurophysiol* 76:109–129.
- Özyürek 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 Cognit Neurosci* 19:605–616.
- Parrish TB, Gitelman DR, LaBar KS, Mesulam M-M (2000): Impact of signal-to-noise on functional MRI. *Magn Reson Med* 44:925–932.
- Peelen MV, Wiggett AJ, Downing PE (2006): Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron* 49:815–822.
- Peigneux P, Van der Linden M, Garraux G, Laureys S, Degueldre C, Aerts J, Del Fiore G, Moonen G, Luxen A, Salmon E (2004): Imaging a cognitive model of apraxia: The neural substrate of gesture-specific cognitive processes. *Hum Brain Mapp* 21:119–142.
- Poldrack RA, Wagner AD, Prull MW, Desmond JE, Glover GH, Gabrieli JDE (1999): Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage* 10:15–35.
- Puce A, Perrett D (2003): Electrophysiology and brain imaging of biological motion. *Philos T Roy Soc B* 358:435–445.
- Pulvermüller F, Huss M, Kherif F, Moscoso del Prado Martin F, Hauk O, Shtyrov Y (2006): Motor cortex maps articulatory features of speech sounds. *Proc Natl Acad Sci USA* 103:7865–7870.
- Riseborough MG (1981): Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *J Nonverbal Behav* 5:172–183.
- Rizzolatti G, Craighero L (2004): The mirror-neuron system. *Annu Rev Neurosci* 27:169–192.
- Rizzolatti G, Fogassi L, Gallese V (2001): Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2:661–670.
- Rodd JM, Davis MH, Johnsrude IS (2005): The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cereb Cortex* 15:1261–1269.
- Romanski LM (2007): Representation and integration of auditory and visual stimuli in the primate ventral lateral prefrontal cortex. *Cereb Cortex* 17(Suppl 1):i61–i69.
- Rosenthal R, Rosnow RL, Rubin DB (2000): Contrasts and Effect Sizes in Behavioral Research: A Correlational Approach. Cambridge: Cambridge University Press.
- Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ (2007): Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb Cortex* 17:1147–1153.
- Saygin AP, Wilson SM, Hagler Jr DJ, Bates E, Sereno MI (2004): Point-light biological motion perception activates human premotor cortex. *J Neurosci* 24:6181–6188.
- Sekiyama K, Kanno I, Miura S, Sugita Y (2003): Auditory-visual speech perception examined by fMRI and PET. *Neurosci Res* 47:277–287.
- 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.
- Seltzer B, Pandya DN (1994): Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. *J Comput Biol* 343:445–463.
- Simos PG, Basile LFH, Papanicolaou AC (1997): Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain Res* 762:29–39.
- Skipper JL, Nusbaum HC, Small SL (2005): Listening to talking faces: Motor cortical activation during speech perception. *NeuroImage* 25:76–89.
- Skipper JL, Nusbaum HC, Small SL (2006): Lending a helping hand to hearing: Another motor theory of speech perception. In: Arbib MA, editor. Action to Language via the Mirror Neuron System. Cambridge: Cambridge University Press. pp 250–285.
- Skipper JL, Goldin-Meadow S, Nusbaum HC, Small SL (2007a): Speech-associated gestures, Broca's area, and the human mirror system. *Brain Lang* 101:260–277.
- Skipper JL, van Wassenhove V, Nusbaum HC, Small SL (2007b): Hearing lips and seeing voices: How cortical areas supporting speech production mediate audiovisual speech perception. *Cereb Cortex* 17:2387–2399.
- Snyder HR, Feigenson K, Thompson-Schill SL (2007): Prefrontal cortical response to conflict during semantic and phonological tasks. *J Cognit Neurosci* 19:761–775.
- Stowe LA, Haverkort M, Zwarts F (2005): Rethinking the neurological basis of language. *Lingua* 115:997–1042.
- 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 Cognit Neurosci* June 25 [Epub ahead of print].
- Swaab TY, Brown CM, Hagoort P (1997): Spoken sentence comprehension in aphasia: Event-related potential evidence for a lexical integration deficit. *J Cognit Neurosci* 9:39–66.
- Thompson LA, Massaro DW (1994): Children's integration of speech and pointing gestures in comprehension. *J Exp Child Psychol* 57:327–354.
- Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ (1997): Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proc Natl Acad Sci USA* 94:14792–14797.
- Thompson JC, Clarke M, Stewart T, Puce A (2005): Configural processing of biological motion in human superior temporal sulcus. *J Neurosci* 25:9059–9066.
- Vaina LM, Solomon J, Chowdhury S, Sinha P, Belliveau JW (2001): Functional neuroanatomy of biological motion perception in humans. *Proc Natl Acad Sci USA* 98:11656–11661.
- Van Petten C, Luka BJ (2006): Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain Lang* 97:279–293.
- Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, Mazoyer B, Tzourio-Mazoyer N (2006): Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage* 30:1414–1432.

- Wagner AD, Pare-Blaevo EJ, Clark J, Poldrack RA (2001): Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron* 31:329–338.
- Wheaton KJ, Thompson JC, Syngeniots A, Abbott DF, Puce A (2004): Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. *Neuroimage* 22:277–288.
- Willems RM, Hagoort P (2007): Neural evidence for the interplay between language, gesture, and action: A review. *Brain Lang* 101:278–289.
- Willems RM, Ozyurek A, Hagoort P (2007): When language meets action: The neural integration of gesture and speech. *Cereb Cortex* 17:2322–2333.
- Wilson SM, Iacoboni M (2006): Neural responses to non-native phonemes varying in producibility: Evidence for the sensorimotor nature of speech perception. *NeuroImage* 33:316–325.
- Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004): Listening to speech activates motor areas involved in speech production. *Nat Neurosci* 7:701–702.
- Wilson SM, Molnar-Szakacs I, Iacoboni M (2008): Beyond superior temporal cortex: Intersubject correlations in narrative speech comprehension. *Cereb Cortex* 18:230–242.
- Wright TM, Pelpfrey KA, Allison T, McKeown MJ, McCarthy G (2003): Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex* 13:1034–1043.
- Wu YC, Coulson S (2005): Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology* 42:654–667.
- Wu YC, Coulson S (2007a): How iconic gestures enhance communication: An ERP study. *Brain Lang* 101:234–245.
- Wu YC, Coulson S (2007b): Iconic gestures prime related concepts: An ERP study. *Psychon B Rev* 14:57–63.
- Zempleni M-Z, Renken R, Hoeks JCJ, Hoogduin JM, Stowe LA (2007): Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *NeuroImage* 34:1270–1279.