Published in final edited form as:

Lang Cogn Neurosci. 2017 ; 32(2): 158–174. doi:10.1080/23273798.2016.1245426.

Multimodal integration of spontaneously produced representational co-speech gestures: an fMRI study

Jill Weisberg^{*},

Laboratory for Language and Cognitive Neuroscience, San Diego State University, 6495 Alvarado Rd., Suite 200, San Diego, CA 92120, USA, 619-594-8069, jweisberg@mail.sdsu.edu

Amy Lynn Hubbard¹, and

Laboratory for Language and Cognitive Neuroscience, San Diego State University, 6495 Alvarado Rd., Suite 200, San Diego, CA 92120, USA, 619-594-8069, amylynnhubbard@gmail.com

Karen Emmorey

Laboratory for Language and Cognitive Neuroscience, San Diego State University, 6495 Alvarado Rd., Suite 200, San Diego, CA 92120, USA, 619-594-8069, kemmorey@mail.sdsu.edu

Abstract

To examine whether more ecologically valid co-speech gesture stimuli elicit brain responses consistent with those found by studies that relied on scripted stimuli, we presented participants with spontaneously produced, meaningful co-speech gesture during fMRI scanning ($n = 28$). Speech presented with gesture (versus either presented alone) elicited heightened activity in bilateral posterior superior temporal, premotor, and inferior frontal regions. Within left temporal and premotor, but not inferior frontal regions, we identified small clusters with superadditive responses, suggesting that these discrete regions support both sensory and semantic integration. In contrast, surrounding areas and the inferior frontal gyrus may support either sensory *or* semantic integration. Reduced activation for speech with gesture in language-related regions indicates allocation of fewer neural resources when meaningful gestures accompany speech. Sign language experience did not affect co-speech gesture activation. Overall, our results indicate that scripted stimuli have minimal confounding influences; however, they may miss subtle superadditive effects.

Keywords

co-speech gesture; fMRI; semantic integration; multisensory integration; multimodal integration

Introduction

To enable the many functions of co-speech gesture, such as directing attention, prosodic cueing, conveying semantic information, and memory enhancement (e.g., Goodwin, 1986;

^{*}Corresponding author: Jill Weisberg, jweisberg@mail.sdsu.edu, phone: 619-594-8069.

¹Current affiliation: Department of Neurology, Psychosomatic Medicine, and Psychiatry, Institute of Health Promotion and Clinical Movement Science, German Sport University Cologne, Am Sportpark Müngersdorf 6, 50933 Köln, (+49) 0152 3627 4844, amylynnhubbard@gmail.com

McNeill 1992; Kelly, Barr, Church, & Lynch, 1999; Morsella & Krauss 2004; and for review, see Kendon 1997), the brain must integrate manual gestures with auditory speech at both the sensory and semantic levels. A persistent challenge in examining which brain regions support these functions has been devising paradigms that present natural, spontaneous co-speech gesture. In order to fit specific experimental criteria, co-speech gesture studies have almost exclusively relied on scripted text with rehearsed (and often contrived) gestures and/or manipulated recordings to create unnatural speech-gesture pairings (Willems, Ozyurek, & Hagoort, 2007; Straube, Green, Bromberger, & Kircher, 2011; Straube, Green, Weis, & Kircher, 2012; Holle, Gunter, Rueschemeyer, Hennenlotter, & Iacoboni, 2008; Holle, Obleser, Rueschemeyer & Gunter, 2010; Green et al., 2009; Kircher et al., 2009; Dick, Goldin-Meadow, Hasson, Skipper, & Small, 2009; Dick, Goldin-Meadow, Solodkin, & Small, 2012). In addition, rather than adopting ecologically valid devices to obscure speech articulators and facial cues (to isolate manual gestures from facial gestures), many studies digitally manipulate video images (e.g., superimposing a filled circle on the face), resulting in an un-natural stimulus appearance which may affect brain activation in key regions. Despite such challenges, findings from these studies largely converge to suggest that the integration of representational gestures with speech relies on a set of regions that includes the posterior superior temporal sulcus (pSTS), parietal cortex, and the inferior frontal gyrus (see Marstaller & Burianova, 2014; Yang, Andric, & Mathew, 2015).

Evidence from various brain imaging techniques and experimental paradigms implicates a region in the human posterior STG/S as a key locus important for multimodal sensory integration, particularly across the visual and auditory modalities (see Calvert, Hansen, Iversen, & Brammer, 2001 for review). STS has been shown to play a role in the sensory integration of visual objects with their associated sounds (Beauchamp, Lee, Argall, & Martin, 2004a), auditory speech with its accompanying mouth movements (Calvert, Campbell, & Brammer, 2000; Campbell et al., 2001), and manual gestures with accompanying speech (Willems et al., 2007; Holle et al., 2008; Hubbard, Wilson, Callan, & Dapretto, 2009).

Semantic integration of gestures with speech may also rely on a region in posterior STG/S. Many gestures, including representational gestures, are not readily interpretable without context afforded by the accompanying speech (McNeill, 1992). Instead, their meaning must be derived based on information contained in speech. Scripted studies of co-speech gesture have exploited this property by manipulating the degree to which representational gestures are congruent with, or disambiguate, the accompanying speech, which in effect, manipulates semantic load (see Ozyurek, 2014, Andric & Small, 2012 for reviews). For example, a sentence might contain the word *grasp*, paired with an iconic gesture indicating sprinkling (Josse, Joseph, Bertasi, & Giraud, 2012), or the word *wrote* with a gesture indicating hit (Willems et al., 2007). These studies report that relative to congruent or unambiguous co-speech gesture, the increased processing effort required to reconcile unnatural, semantically incongruous (or ambiguous) speech-gesture pairings, elicits increased activity in pSTS/G. Such evidence suggests that the posterior temporal lobe participates in the semantic integration of the audio speech signal with visual gestures.

Similarly, scripted studies show that semantic processing demands modulate activity in frontal regions. For example, the left inferior frontal gyrus (IFG; and sometimes the right IFG) shows increased activation during perception of semantic gesture/speech incongruencies (Willems et al., 2007; Straube et al., 2013; Green et al., 2009; Dick et al., 2009; 2012). Stronger responses in IFG are also found for speech with gesture than for speech alone (Dick et al., 2009); for meaningful than meaningless gestures (Holle et al., 2008); for speech with metaphoric gestures than with iconic gestures (Straube et al., 2011); and for speech with incongruent than with congruent pantomime (Willems, Ozyurek, & Hagoort, 2009). Although these findings consistently suggest a role for inferior frontal cortex in the semantic integration of gesture and speech, others suggest that detection and resolution of semantic incongruencies might better explain its role in co-speech gesture perception (e.g., Noppeney, Josephs, Hocking, Price, & Friston, 2008).

Inferior parietal cortex is also implicated in the semantic integration of representational gestures with speech, as evidenced by increased left intraparietal sulcus (IPS) activation for mismatched, relative to matched, gesture with speech (Willems et al., 2007). Again, such activation is thought to reflect increased semantic load for the less natural and more ambiguous pairings: decoding unrelated gestures with speech requires more effort than decoding related gestures, leading to greater activation during integration. However, a study comparing responses for speech with meaningful vs. grooming gestures found greater bilateral parietal activation for the meaningful gestures (Holle et al., 2008). One possibility is that such activity reflects increased simulation costs during observation-execution mapping. In this case, increased simulations may be needed to integrate potentially meaningful gestures, compared to meaningless grooming movements in the context of the speech signal (Green et al., 2009).

Across studies, various approaches have been used to identify multisensory integration regions (and multisensory neurons). One approach often associated with animal electrophysiological studies identifies regions where unimodal responses are greater than the baseline *and* the response to a multimodal stimulus exceeds the sum of the unimodal responses (e.g. Stein, & Meredith, 1993; Laurienti, Perrault, Stanford, Wallace, & Stein, 2005). This approach, referred to as superadditivity, was successfully used in human MRI studies of visual-auditory speech integration (Calvert et al., 2000, 2001). This approach is a stringent test of multisensory integration, in that it eliminates regions where either unimodal stimulus fails to elicit a significant positive response. As such, the superadditivity approach ensures that the integration measure is not an artifact of a negative unimodal response. Calvert et al. (2000) used an even stricter approach, requiring that a region show a superadditive response to congruent audiovisual speech *and* a subadditive response to incongruent stimuli in order to qualify as a multimodal integration region. In that study, the only region meeting this criteria, which parallels the response pattern of multisensory integrative neurons in non-human primates, was located in the ventral bank of left STS.

Some argue that superadditivity is too stringent a measure, given the small changes detected by BOLD, and propose other criterion measures. For example, one alternative criterion for multimodal integration is that the multimodal response exceeds the mean of the unimodal

responses (mean method), and another requires the multimodal response to exceed the maximum of the unimodal responses (max method; Beauchamp 2005).

A goal of the present study was to determine whether findings from scripted representational co-speech gesture studies could be replicated for more ecologically valid stimuli. That is, do the same frontal, temporal, and parietal regions associated with the integration of scripted, manipulated, or rehearsed representational co-speech gesture support the integration of gestures and speech when both are spontaneous, natural, and unrehearsed? To our knowledge, no prior neuroimaging study has examined brain responses to representational co-speech gesture in this manner. However, Hubbard et al. (2009) used such an approach to explore the integration of manual beat gestures (i.e. meaningless rhythmic gestures) accompanying speech (see also Hubbard et al., 2012). Here, we use the same superadditivity measure as Hubbard et al. (2009) to allow direct comparison of results for implementations of two different categories of naturally-produced co-speech gesture, i.e., beat gestures and representational gestures.

An innovative feature employed by Hubbard et al. (2009) was the naturalistic setting created by filming a speaker standing in a kitchen, facing the camera, her legs occluded by the stove and countertop, and her head occluded by the cupboards above the stove to ensure that gesture was the only speech-accompanying movement visible (i.e. no mouth or head movements were visible). Second, the freely moving female speaker *extemporaneously* answered questions about her life and experiences, conversing with an off camera interlocutor in an adjoining room. From a two-hour recording, Hubbard et al. (2009) culled short video segments (~18s) containing beat and rhythmic co-speech gestures (i.e. excluding segments with representational gestures). Beat gestures are mono- or bi-phasic, repetitive movements serving a rhythmic function linked to speech prosody, with minimal semantic content. Participants passively viewed these video segments during fMRI scanning under unimodal and bimodal conditions: segments were presented as speech with beat gestures (bimodal co-speech gesture), as video without audio (unimodal beat gesture perception), and as audio with the speaker standing motionless (unimodal speech perception). Control conditions presented speech with manual signs from American Sign Language (ASL) that had no meaning for the participants.

Hubbard et al. (2009) identified integration effects in left anterior and right posterior STG/S for beat gestures accompanying speech. In left anterior STG/S, increased activation for speech accompanied by beat gestures, compared to speech with meaningless movements, suggested sensitivity of anterior STG/S to the integration of synchronous beat gestures with speech, perhaps boosting comprehension by focusing attention on prosody. Importantly, only the more posterior region of right STG/S (incorporating the planum temporale) showed a superadditive response to speech accompanied by beat gestures. That is, in right planum temporale the response to speech with gesture exceeded the summed responses to speech without gesture and gesture without speech. However, responses in this region did not differentiate between speech accompanied by beat gesture vs. meaningless control movements, as did the more anterior region, suggesting that posterior STG/S is sensitive to the potential of multisensory stimuli to convey meaning, particularly via prosody.

Superadditive integration effects such as those just described had not previously been reported for speech-gesture integration, but it is not known whether earlier studies tested for superadditivity. Instead, most previous studies of co-speech gesture focused on meaningful, representational co-speech gestures, which are largely ambiguous without the context from speech (yet still meaningful, and less ambiguous than beat gestures). Therefore, it is unknown whether the integration effects found by Hubbard et al. (2009) for naturalistic prosodically-linked co-speech beat gestures will generalize to perception of similarly naturalistic, representational co-speech gestures. The current study explores this question using short video segments culled from Hubbard et al.'s (2009) original two-hour recording. We selected only segments that contained representational co-speech gestures (as rated by a separate group of participants, see Methods) and presented them under four conditions: speech (audio only) with a motionless speaker (S); co-speech gestures without audio (G); co-speech gestures with concurrent speech (SG); and still frames of the speaker at rest, without audio (Still baseline). If cortical responses to spontaneous, representational, audiovisual co-speech gestures pattern like scripted co-speech gesture, we expected integration effects in frontal, posterior temporal, and parietal regions. However, if scripted stimuli unintentionally introduce confounding effects (e.g. un-natural manipulations may cause visual attention effects), we may find that some previously reported activations are epiphenomenal.

In addition, based on Hubbard et al.'s spontaneous co-speech beat gesture findings (Hubbard et al., 2009), we expect that superadditive responses might be revealed, particularly in posterior temporal cortex, and that the semantic content of representational gestures may shift the locus of integration from sensory to higher-order cortical regions.

Prior studies in our laboratory demonstrated that sign language experience can impact co-speech gesture production (e.g. increased proportion of iconic gestures; greater handshape diversity; Casey & Emmorey, 2009; Casey, Larrabee, & Emmorey, 2012), but how ASL knowledge effects the co-speech gesture perception system has yet to be examined. We therefore included a group of hearing bimodal bilinguals, with native fluency in American Sign Language (ASL) and English, and compared their responses to co-speech gesture with those of monolingual speakers. We hypothesized that routine use of a visual-manual language system for semantically meaningful communication might influence sensitivity to the communicative functions of co-speech gesture. Compared to monolingual participants, we hypothesized that representational co-speech gestures might differentially activate the brains of hearing signers in regions associated with biological motion perception (e.g. posterior STG) and action understanding (prefrontal and parietal cortices), key nodes in the network for comprehension of visual-manual languages (e.g. Söderfeldt et al., 1997; MacSweeney et al., 2002; Emmorey, McCullough, Mehta, & Grabowski, 2014; Weisberg, McCullough, & Emmorey, 2015).

Methods

Participants

Fourteen English monolinguals (7 females, mean age = 25.1 years, SD = 5.05; mean education = 15.5 years, SD = 1.83) and 14 hearing ASL-English bilinguals (7 female, mean

age = 22.4 years, SD = 3.99, mean education = 14 years, SD = 1.95) participated in the experiment. Data from three additional monolingual participants were excluded from analysis due to excessive movement during scanning. The bilingual participants were all “children of deaf adults” (CODAs) born into deaf signing families. All participants were right-handed, healthy, and reported no neurological, behavioral, or learning disorders. Informed consent was obtained according to procedures approved by the UCSD and SDSU Human Research Protection Programs.

Materials & Procedure

Stimuli consisted of video clips culled from two hours of natural (i.e. unscripted) co-speech gesture produced by a female native English speaker, naïve to the purpose of the recording. Details of the original video recording can be found in Hubbard et al., 2009 and will be described only briefly here. The speaker answered questions about her hobbies and experiences while standing in her own kitchen. A thin board was affixed to the wall in front of the speaker such that it appeared to be a cupboard obscuring the speaker’s face, and a stove built in to the kitchen counter obscured the speaker’s legs. Hence, in the video recording, the speaker was visible only from the lower neck to her hips, occluding head movement, facial expressions, visual speech, and other communicative cues, and ensuring that gesture was the only perceptible speech-accompanying movement. Importantly, the configuration allowed the speaker to maintain eye contact with her interlocutor, positioned in an adjoining room across the counter, throughout the recording. A separate recording depicted the speaker standing motionless in the same configuration.

After importing the video to a Macintosh computer, 22 audiovisual segments of continuous speech containing representational gestures were exported using Final Cut Pro 6.06. The segments (duration range = 14–19s) were cut from five different points in the two-hour recording, representing five different conversational themes (3–6 segments per theme). We based final stimulus selection on ratings from a separate group of 20 participants who did not participate in the imaging study. Raters viewed each audiovisual segment twice and then twice viewed shorter clips of each individual gesture (with audio) edited from the longer segments. Each rater viewed 80–120 individual gestures from two of the five themes, with theme counterbalanced across raters. For each individual audiovisual gesture clip, participants indicated whether they felt the gesture a) might have some meaning when combined with speech; b) is only rhythmic; c) is not a gesture. They also rated the clarity of each gesture (scale of 1–5 where 1 = unclear and 5 = perfectly clear) and reported what they thought each gesture might represent. For example, an isolated co-speech gesture (two hands form a loose sphere) produced with the phrase “a lot of people who came out” was rated as *meaningful* by a majority of participants; clarity ratings for the gesture ranged from *mostly* to *perfectly clear*, and representations from three participants included “a lot of people”; “more than one person”; and “a lot.” We included 18 segments in the imaging study based on selection criteria that required at least two meaningful gestures per segment with > 75% agreement across raters that those gestures were meaningful. Four segments did not meet those criteria and thus, were excluded from the imaging experiment.

The 18 selected segments ranged in length from 14s to 19s, contained between 2–7 meaningful gestures each, and represented all five conversational themes. The range of segments per theme was 1–6, as some themes occupied more time than others in the original two-hour recording. We randomly divided the 18 segments into three groups of six, assigned each group of six to a specific condition (S, G, SG), and reassigned each group to a different condition for each of the three imaging runs. Runs were constructed such that participants viewed all 18 segments once per imaging run (three times across the experiment), and each segment appeared in a different condition (S, G, or SG) in each of the three runs. Each participant underwent three consecutive eight-minute scans (for a total imaging time of approximately 24 minutes). During each scan, a participant viewed six stimulus cycles, each consisting of one segment per experimental condition (S, G, SG) plus a baseline condition segment (still frames of the speaker at rest without no audio; ‘Still’) (Figure 1). Presentation order of the four conditions was kept constant across cycles within a run, but was pseudo-randomized across runs to prevent item and order effects. Each condition appeared an equal number of times in each position, fully counter-balanced for all segments across all runs, and order of runs was counterbalanced across participants.

During scanning, participants passively viewed each video, and we instructed them to watch and listen to each clip carefully, informing them that they would be asked to answer one true/false question at the completion of each run and to complete a short true/false questionnaire immediately following the scanning session. Prior to scanning, all participants received training with video segments not used during scanning. During training we emphasized that the questions were very simple. We explained that we would not try to trick them or trip them up, and that the questions were not detailed or difficult, but were designed simply to ensure that they attended to the stimuli during scanning. An example of a question after a scanning run is “True or false: The speaker likely knows how to surf;” and a sample question from the post-scan questionnaire is “The speaker prefers to teach a) adults; b) girls; c) dogs; or d) boys.” Mean accuracy on the post-scan test was 94.29% (SD = 5.12) for bimodal bilinguals and 95.59% (SD = 5.56) for monolinguals, indicating that all participants attended to the stimuli throughout the entire scanning session.

An XGA video projector connected to a MacBook Pro computer running Quicktime projected the stimuli onto a screen at the foot of the scanner bed, viewed via a mirror positioned atop the head coil.

Data Acquisition and Analysis

MRI data were acquired at the Center for Functional MRI at the University of California, San Diego on a 3-Tesla GE Signa Excite scanner equipped with an eight-channel phased-array head coil. We acquired high-resolution anatomical scans for each participant for use in spatial normalization and anatomical reference (T1-weighted FSPGR, 176 sagittal slices, FOV = 256mm, matrix = 256×256 , 1mm^3 voxels). Functional data were acquired using gradient-echo echo-planar imaging (40 T2*-weighted EPI axial slices interleaved from inferior to superior, FOV = 224mm, 64×64 matrix, 3.5mm^3 voxels, TR = 2000ms, TE = 30ms, flip angle = 90°). Each of three functional scans lasted eight minutes (260 whole-

brain volumes). Six additional volumes at the start of each scan were removed during preprocessing.

All MRI data were processed and analyzed using AFNI (Cox 1996). Individuals' T1 anatomical images were spatially normalized to the stereotactic space of Talairach and Tournoux (1998) using a 12-parameter affine registration algorithm implemented in AFNI's @auto_tlrc program with the TT_N27 template as the reference (Colin27). We report all coordinates in that space. For each individual we applied this transformation to their T2* functional data and at the same time corrected head movements with 3dvolreg using the second volume of the first run (pre-steady state) as the reference. The output was smoothed with a 7mm Gaussian filter and scaled to a mean of 100 to derive percent signal change. Prior to statistical analysis, any timepoints where the Euclidean norm of the derivatives exceeded 1 were censored.

Using 3dDeconvolve we generated an $\times$ matrix for multiple regression analysis, which we implemented using 3dREMLfit to improve the accuracy of parameter and variance estimates (Chen, Saad, Nath, Beauchamp, & Cox, 2012). Nuisance regressors in these analyses included the six movement parameters output during motion correction and a Legendre polynomial set ranging from zero to third order for each run to account for slow drifts. Three regressors of interest included one for each experimental condition (S, G, SG) modeled using AFNI's dmbLOCK basis function with variable durations across blocks. The still video baseline blocks comprised the baseline for the regression model and thus were not modeled with a regressor.

To examine group-level effects we entered individuals' voxel-wise response estimates (beta-weights) and their corresponding t-values for each stimulus type and/or contrast of interest in a second-level, mixed effects meta-analysis (3dMEMA; Chen et al., 2012). As our analyses identified no significant differences between groups, we report results collapsed across the larger group of 28 participants, with mean age of 23.8 years ($SD = 4.69$) and mean education of 14.6 years ($SD = 1.93$). All reported results exceeded a false discovery rate correction for multiple comparisons (designated as q) thresholded at $q < .05$, unless otherwise noted. We report all cortical clusters of 10 or more contiguous voxels. We conducted pairwise contrasts between all experimental conditions (S vs. G; S vs. SG, G vs. SG).

In order to discover regions that might respond uniquely to bimodal input, we searched for regions more strongly activated by speech with gesture vs. each unimodal condition separately (SG vs. S, SG vs. G), regardless of whether the region showed significant activation for unimodal input. To examine how neural responses to speech were modulated in the presence of gesture and how responses to gesture were modulated in the presence of speech ($SG > S$, $SG > G$), we again searched for regions more strongly activated by speech with gesture versus each unimodal condition. We report all clusters that showed a positive response to either unimodal stimulus and where speech with gesture was significantly different from gesture alone, or from speech alone. That is, significant clusters where the unimodal condition showed a negative response or no response at all were not included in this report. Lastly, following Hubbard et al. (2009), we identified regions associated with

speech-gesture integration using a superadditivity measure (see Beauchamp, 2005; Calvert et al., 2000; Campbell et al., 2001; Callan et al., 2004; Hubbard et al., 2009, 2013). Specifically, this contrast identified voxels where $SG > (S + G)$. All clusters that show a superadditive response and a positive response to both unimodal conditions are reported here. This report does not include regions that appear to be superadditive simply because a unimodal stimulus elicited either no response or a negative fMRI response.

Results

As noted, we identified no differences between the ASL-English bilingual and monolingual groups, and we therefore report results for data collapsed across all participants. All results exceeding a statistical threshold corrected for multiple comparisons at $q < .05$, focusing first on pairwise contrasts between experimental conditions (speech, gesture, speech with gesture) and then on regions where superadditive responses indicated integrated processing of speech and gesture (responses to the bimodal condition exceeded the sum of responses to the unimodal conditions).

Speech alone vs. Gesture alone

The group-level analysis directly contrasting brain responses to speech with responses to gesture yielded extremely robust modality-related activation differences, requiring implementation of a high statistical threshold ($q < 10^{-5}$). Speech elicited greater activation than gesture primarily in regions associated with auditory processing, while gesture elicited greater activation in regions associated with visual processing (Figure 2A, Table 1). Specifically, large clusters showing preferential responses to speech were focused in auditory cortex (superior temporal gyrus, bilaterally), and extended anteriorly toward the temporal pole (Table 1 about here). In contrast, gesture elicited greater activation than speech in widespread bilateral occipitotemporal visual areas (occipital cortices, extending into the fusiform gyrus in the right hemisphere) and bilateral parietal cortices (superior parietal lobule, extending into the inferior parietal lobule and the supramarginal gyrus).

In order to examine regions common to speech perception and gesture perception, we conducted a conjunction analysis by separately contrasting each condition with the still baseline condition at a lower, but still stringent threshold ($q < .05$) and overlaying the resultant statistical maps. This analysis revealed that activation for speech perception and gesture perception overlapped in several regions, bilaterally, including a large swath of occipital cortex, superior temporal sulcus, inferior frontal and precentral gyri, medial frontal cortex, and left supramarginal gyrus (Figure 2B). Thus, although the direct contrast revealed robust modality-related differences, audiovisual speech and visual gesture perception share overlapping regions of activation associated with visual sensory perception and language.

Unimodal (S or G) versus Bimodal (SG)

We approached this analysis in several different ways. First, we searched the brain for regions more responsive to bimodal speech with gesture than to speech alone, and separately, more to the bimodal condition than to gesture alone. This analysis was agnostic as to whether or not a region responded significantly more to the unimodal stimuli than the

Still baseline in order to identify regions uniquely activated by the combination of speech with gesture (but not active for either input alone). The results ($q < 10^{-5}$) indicated that compared to speech alone (S), all regions more active for speech with gesture (SG) overlapped with regions associated with gesture (G) perception (i.e. $G > \text{baseline}$), and conversely, all regions more active for SG than for gesture alone overlapped with those associated with speech perception (i.e. $S > \text{Still}$). Thus, the whole brain analyses comparing bimodal with unimodal input revealed no regions uniquely active for the integrated perception of speech with gesture that weren't also responsive during unimodal perception of either speech or gesture.

Second, we conducted a similar analysis confined to regions that showed significant responses during unimodal speech or gesture perception (i.e. G vs. SG in regions where $G > \text{Still}$, and S vs. SG in regions where $S > \text{Still}$, both $q < .05$). This analysis explored whether activation in gesture- and/or speech-responsive regions is modulated by bimodal perception of speech combined with gesture. Compared to speech alone, speech with gesture showed reduced activation in bilateral insula, and increased activation in bilateral inferior frontal gyrus (pars triangularis), superior frontal gyrus, precentral gyrus, STG/S, and lateral occipital gyrus (BA 17) (Table 2). Compared to gesture alone, speech with gesture showed decreased responses in right superior parietal cortex and middle temporal gyrus, and increased activation in bilateral inferior frontal gyrus (pars triangularis), precentral gyrus, STS, posterior occipitotemporal cortex (BA 17/18 and the fusiform gyrus) and the left supramarginal gyrus (Table 2 about here). Inspection of Table 2 reveals that bimodal perception of speech with gesture resulted in highly similar regions of increased activity whether compared to gesture alone or to speech alone. That is, in bilateral IFG, precentral gyrus, STS, and lateral occipital gyrus, speech combined with representational gestures elicits greater activation than unimodal speech or gesture input. These regions also overlap with those identified by the conjunction analysis (Figure 2B) as significantly more active for speech vs. the baseline condition *and* for gesture vs. the baseline condition.

Superadditivity: (SG) > (S + G)

Lastly, in order to allow direct comparison with the results of Hubbard et al., (2009), who investigated co-speech beat gestures using a similar experimental design, we searched for regions associated with the integration of speech with meaningful gestures using the same established superadditivity measure. This measure specifies two criteria: 1) within the cluster, the multimodal stimulus response exceeds the summed responses of the unimodal stimuli; and 2) both unimodal stimuli elicit significant responses within the cluster. Specifically, we examined the data for regions where speech with gesture elicited responses significantly greater than the sum of the responses to speech alone and gesture alone ($q < .05$), and report all clusters where the unimodal stimuli elicited significant responses.

Two small clusters meeting both superadditivity criteria emerged from this analysis, one in left posterior STS and the other in left precentral gyrus. In both of these left hemisphere clusters, responses to speech with gesture significantly exceeded the sum of the responses to gesture and to speech alone ($q < .05$; Figure 3, Table 3). In contrast, bilateral occipital

cortex, in and around primary visual cortex, was the only area where the summed unimodal responses exceeded activation for speech with gesture (Table 3 about here).

Discussion

We examined brain regions supporting multimodal integration in sensory and semantic domains by comparing brain activation to video clips of extemporaneous speech accompanied by spontaneous, meaningful gestures, to activation during perception of unimodal speech and unimodal gesture. Compared to speech and to gesture alone, perception of meaningful gestures with speech led to heightened activity in several regions, most notably in bilateral posterior STG/S, precentral gyrus, and inferior frontal gyrus, suggesting that these regions play a role in multimodal integration at the sensory and/or the semantic level. Furthermore, a subset of the voxels within the left hemisphere STG/S and premotor regions showed a superadditive response for speech with gesture. That is, within these small, circumscribed regions, responses to speech with gesture exceeded the sum of responses to the unimodal conditions. We suggest that these discrete regions in posterior temporal and premotor cortices may participate in both sensory- and semantic-level integration of meaningful co-speech gestures. That is, in the service of integrating information across sensory and semantic inputs, perception of speech accompanied by manual gestures that convey complementary semantic information boosts responses in these regions beyond what would be expected by simply summing the unimodal responses.

Multimodal integration of co-speech gesture in posterior STG/S

The large bilateral STG/S clusters showing heightened responses for speech with gesture were centered in the auditory association cortex of each hemisphere. Their locations coincide with the parts of STG/S that the conjunction analysis revealed to be active for speech alone and for gesture alone (versus the baseline condition). Thus, the larger responses for speech with gesture are not an artifact of the analysis caused by a lack of activation to one or both unimodal conditions. In addition, STG/S responded more to speech with gesture than to unimodal speech *and* to unimodal gesture. This pattern of results strongly suggests that these regions in bilateral STG/S participate in multimodal integration of co-speech gesture.

Audiovisual integration-related effects in STG/S previously demonstrated in both animal and human studies are consistent with the idea that multimodal integration relies, in part, on posterior temporal cortex. In monkeys, electrophysiological recordings identified discrete patches of neurons in STS that responded equally to visual and auditory stimuli, in contrast to patches that responded only to one modality (Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Dahl, Logothetis, & Kayser, 2009; Seltzer & Pandya, 1994). Similarly, in humans, high-resolution brain imaging identified multisensory patches of neurons in STS that responded to unimodal auditory and visual stimuli, and also showed an audiovisual response that exceed both unimodal responses (Beauchamp, Argall, Bodurka, Duyn, & Martin 2004b). The multisensory patches were distinct from patches that preferentially responded only to auditory or only to visual stimuli; in the unimodal patches, audiovisual stimuli elicited responses no greater than that of a patch's preferred modality. Though imaging

studies conducted at more typical resolutions cannot detect such patches, they too consistently demonstrate that an area in posterior STG/S plays a key role in sensory integration across diverse stimuli and tasks (e.g. multisensory object recognition, audiovisual speech comprehension; Calvert et al., 2000, 2001; Beauchamp et al., 2004a; van Atteveldt, Formisano, Goebel, & Blomert, 2004; Szyck, Stadler, Tempelmann, & Muentel, 2012).

However, another possibility is that the region we identified in posterior superior temporal cortex supports semantic integration of gestures with speech. Consistent with this idea, posterior temporal cortex has been linked to semantic integration by co-speech gesture studies that varied semantic load by manipulating the meaningfulness of manual gestures (e.g., congruent metaphorical, iconic, or emblematic gestures versus incongruent, grooming, or nonsense gestures; for review, see Ozyurek, 2014; Andric & Small, 2012). Resolving semantic disparities during perception of speech with incongruent, compared to congruent gestures, is thought to increase semantic processing demand, leading to heightened activity in posterior temporal cortex during semantic integration (Green et al., 2009). Similar congruency effects occur for audio-visual speech integration (i.e. auditory speech with visual mouth movements), as demonstrated by fMRI, MEG, and EEG studies (Calvert et al., 2000; van Atteveldt, Blau, Blomert, & Goebel, 2010; Raij, Uutela, & Hari, 2000; Callan, Callan, Kroos, & Vatikiotis-Bateson, 2001). In non-human primates, electrophysiological studies of action-sound integration also implicate posterior temporal cortex in integration of meaning across the auditory and visual modalities (e.g. Barraclough et al., 2005; Dahl et al., 2009).

Although the averaging of MRI responses over voxels that each contain millions of neurons prohibits distinction between whether the increased activation we observed in STG/S reflects sensory or semantic integration, this region is known to be heterogeneous with respect to its function in different sectors and in different functional contexts (e.g., Scott & Johnsrude, 2003). Therefore, some sectors may support multimodal sensory perception while others may support multimodal semantic perception.

Within the large left hemisphere STG/S cluster where speech with gesture elicited larger responses than the unimodal conditions, we identified a smaller cluster with a superadditive response for speech with gesture (i.e., speech with gesture > speech alone + gesture alone summed). The peak of this smaller cluster was nearly identical to that of the larger STG/S cluster in the left hemisphere. The smaller cluster appears to be functionally distinct from the bordering area that shows a heightened, but not superadditive response for speech with gesture. We hypothesize that in communicative contexts, multiple levels of integration may converge in this superadditive region of left STG/S to support comprehension of the combined audio-visual signal conveyed by vocal speech and manual gestures. That is, the circumscribed cluster where the multimodal response exceeds the sum of its parts may participate in sensory *and* semantic integration, whereas the surrounding area participates in sensory *or* semantic integration.

Consistent with this hypothesis, superadditive responses have been identified in multisensory neurons located in both cortical and subcortical regions by electrophysiological animal studies (e.g. Barraclough et al., 2005; Meredith & Stein, 1986), and in human brain imaging studies of multisensory integration (Gottfried & Dolan, 2003; Stevenson,

Geoghegan, & James, 2007; Calvert et al., 2001; Werner & Noppeney, 2010). For example, Werner and Noppeney (2010) found that superadditive activation in STS predicted semantic categorization of audiovisual, compared to unimodal depictions of musical instruments and tools. Similarly, Gottfried and Dolan (2003) demonstrated superadditive responses in orbital frontal cortex and STS, only for congruent, but not for incongruent olfactory and visual stimuli. In the language domain, Calvert et al. (2000) demonstrated that superadditive responses in STS depended on semantic congruence of audiovisual speech.

Few imaging studies of meaningful co-speech gesture perception were designed to detect superadditive responses (i.e. inclusion of two unimodal conditions to test their sum against a multimodal condition). However, Hubbard et al. (2009) investigated integration of speech with co-speech beat gestures by contrasting activation for speech with spontaneous manual beat gestures to the summed responses for speech alone and beat gestures alone. The stimuli used by Hubbard et al. and the current study were extracted from the same recording of spontaneous speech with gestures, with video segment selection criteria differing across the two studies only by gesture type: Hubbard et al. selected segments with beat gestures; we selected segments with representational gestures. Hubbard et al. identified a superadditive effect for speech with beat gestures focused in right posterior planum temporale, anterior and contralateral to the left STG/S region where we observed a superadditive effect for representational gestures (Hubbard et al. Talairach coordinates² $x = 57$, $y = -26$, $z = 9$). They suggested that as part of a communicative act, rhythmic beat gestures, though non-meaningful, alter the processing of speech in a region associated with prosody and melody (i.e. planum temporale; Hubbard et al., 2009). The superadditive locus for representational gestures in the left hemisphere is consistent with left hemisphere dominance for semantic linguistic processing, while the right locus in Hubbard et al. is consistent with the right hemisphere's role in rhythmic, prosodic processes (Witteman, van Ijzendoorn, van de Velde, van Heuven, & Schiller, 2011).

Multimodal integration of co-speech gesture in premotor cortex

We found the same response pattern in premotor cortex as we observed in posterior superior temporal cortex. Specifically, in bilateral premotor cortex, activation increased for spontaneous representational gestures with speech, relative to unimodal speech or gestures. Although we did not directly manipulate stimulus semantics, speech accompanied by representational gestures conveys more semantic information than speech or gestures perceived in isolation. Thus, we hypothesize that the increased activation in premotor cortex likely reflects semantic associations between the auditory speech and visual gesture inputs. Compatible with this notion, neurons identified in monkey ventral premotor cortex (area F5) respond when a monkey performs a specific action (e.g. grasping), observes the action, views a graspable object in the absence of action, or hears a sound associated with the action (Murata et al., 1997; Rizzolatti et al., 1988; Kohler et al, 2002). Kohler et al. (2002) found that some of those neurons fired more intensely for actions accompanied by their associated sounds than for presentation of the action or sound alone, but showed no response to unnatural, contrived sound-action pairs. They proposed that this response pattern, coupled

²Lancaster transformation from Montreal Neurological Institute (MNI) coordinates (Lancaster et al., 2007; Laird et al., 2010).

with the fact that the audiovisual neurons also responded during action execution, indicates that these neurons may support conceptual representations of actions in addition to more basic sensorimotor information.

In humans, premotor cortex (along with parietal cortex and other regions) is part of a network engaged during action execution and observation and is thought to participate in simulations to aid action comprehension (Buccino et al., 2001; Jeannerod, 2001). Neuroimaging studies have demonstrated that many tasks involving observation of body actions or of manipulable objects with motor affordances (e.g. passive viewing, naming, matching, imagining) activate premotor regions overlapping with those that respond during movement execution. For example, Buccino et al (2001) showed that during observation of hand, mouth, and foot actions, premotor cortex engages in a somatotopic manner that mirrors that of movement execution (see also Hauk, Johnsrude, & Pulvermüller, 2004). In keeping with a semantic processing role for premotor cortex, simulations aiding action comprehension, combined with the language system, are thought to form part of a foundation for conceptual knowledge about actions and objects we encounter in our environment (Barsalou, 2009; Martin, 2009; Grafton, Fadiga, Arbib, & Rizzolatti, 1997).

One consequence of simulations involving premotor cortex may be top-down modulation of sensorimotor processing, contingent on the semantic similarity between speech and accompanying gestures. For example, in an fMRI adaptation study by Josse, Joseph, Bertasi, and Giraud (2012), participants decided whether an auditory test stimulus contained a word from a preceding audiovisual clip of speech with congruent or with incongruent gestures, or of speech alone. The test stimulus was either novel or repeated from the previous audiovisual presentation. Compared to speech alone, greater repetition suppression occurred in dorsal premotor cortex when test stimuli followed speech with congruent gestures. In contrast, repetition enhancement, rather than suppression, occurred when test stimuli followed speech with incongruent gestures. Josse et al. proposed that cortical representations overlap more extensively when words and gestures are congruent and less extensively when incongruent. The extent of overlap, in turn, determines whether a smaller or larger pool of neurons respond, leading to repetition suppression or enhancement, respectively. In a separate study, Skipper, Goldin-Meadow, Nusbaum, and Small (2009) demonstrated that the type of movement that accompanies speech could modulate functional connectivity between premotor cortex and other cortical regions. In that study, perception of speech with meaningful (but not meaningless) gestures yielded the strongest connectivity between premotor cortex and semantic processing regions (e.g. anterior STG/S, supramarginal gyrus); in contrast, speech with mouth movements (no gesture) resulted in strong connectivity between premotor cortex and regions involved in motor and phonological processes (e.g. posterior STG/S; Skipper, Nusbaum, & Small, 2005, 2009; Calvert et al., 2000). Thus, the function of premotor cortex appears to change depending on context, working closely with posterior superior temporal areas to support semantic decoding when meaningful manual gestures accompany speech, and with more anterior motor regions to support phonological processing when facial movements accompany speech. Taken together, the evidence discussed above converges to indicate a role for premotor cortex in conceptual processing, requiring the integration of sensorimotor information with semantic processing of co-speech gesture.

Within the left premotor cluster that showed an increased response for speech with gesture compared to either modality alone, we identified a small cluster of voxels exhibiting a superadditive response, analogous to that detected within STG/S. Willems et al. (2009) demonstrated a superadditive effect focused more inferiorly, in left IFG, during integration of speech with metaphoric gestures, compared to unimodal inputs. Those semantic congruency effects occurred in left IFG for the bimodal condition whether speech was accompanied by iconic co-speech gestures or action pantomimes; but in bilateral posterior temporal cortex (pSTS/MTG), the superadditive response occurred only for speech with pantomime. They suggested that IFG and posterior temporal cortex serve different roles for semantic integration of action with language, with pSTS/MTG integrating information about stable, common representations (such as pantomime), and IFG integrating information to form novel associations (such as gestures that are ambiguous without speech). In the current study, the co-speech gesture stimuli may have generated novel associations — even meaningful gestures are not conventionalized and are difficult to comprehend without context provided by speech. However, unlike Willems et al. (2009), we did not observe qualitatively different responses in frontal, compared to posterior temporal cortex, and the foci of the clusters in Willems et al. (2009) were located inferior (left IFG) and lateral (pSTS/MTG) to ours. The different findings between these studies may be related, in part, to stimulus and paradigm choices. Nevertheless, the similarity of response patterns we found in frontal and temporal cortices suggest that like in STG/S, a small, circumscribed region in the left precentral gyrus may play a dual role, supporting both semantic and sensory integration.

Anatomical and functional imaging studies support the idea that parts of premotor cortex serve a dual role in semantic and sensory processing. As part of the dorsal route for speech perception proposed by Hickok and Poeppel (2007), overlapping representations of the manual and speech articulators (e.g. Meister, Buelte, Staedtgen, Boroojerdi, & Sparing, 2009) may enable dorsal premotor cortex to participate in audiovisual and sensorimotor integration (Josse et al. 2012). For example, premotor cortex, along with the planum temporale and STG, are thought to form a network that supports sensorimotor integration for error detection and correction by providing auditory feedback during speech production (Hickok, Houde, & Rong, 2011; Tourville, Reilly, & Guenther, 2008). If dorsal premotor cortex participates in integration across various systems, the superadditive response may be one marker of integration across multiple domains. In the context of the present study, given the documented role of premotor cortex in conceptual knowledge (Pulvermüller et al., 2010; Martin et al., 2009; Weisberg, van Turenhout, & Martin, 2007; Willems, Hagoort, & Casasanto, 2010), superadditive premotor responses may reflect both sensory (or sensorimotor) and semantic integration in the service of multimodal language comprehension.

Multimodal integration of co-speech gesture in IFG

Similar to our findings in STG/S and precentral gyrus, the inferior frontal gyrus showed an increased response, bilaterally, for speech with gesture, relative to perception of speech or gesture alone. However, in contrast to STG/S and precentral gyrus, we found no indication of superadditive responses within the IFG of either hemisphere. This suggests that this part of the IFG is more uniform in function, integrating information across only one domain, and

unlike the STG/S and precentral gyrus, does not encompass a discrete region that participates in integrating information across both the sensory and semantic domains. That is, during comprehension of speech with meaningful gestures, IFG is recruited to support either sensory integration of auditory and visual stimuli, or the semantic information conveyed by the speech and gesture streams. In this case, the more plausible candidate is semantic integration, as the inferior frontal gyrus is closely associated with many language functions, and in particular with semantic processing (Bookheimer, 2002; Thompson-Schill, 2003).

A variety of experimental paradigms that manipulated semantic load have demonstrated that IFG participates in speech-gesture integration. The inferior frontal gyrus is more strongly activated for speech-gesture combinations with greater semantic load than those with lower semantic load (e.g. congruent versus incongruent gestures; Green et al., 2009; Dick et al., 2009; Straube et al., 2011), presumably because greater effort is required to integrate the disparate messages conveyed by the words and manual gestures (Dick et al., 2009; Willems et al., 2009). However, as inferior frontal cortex is also associated with selection between competing alternatives (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Moss et al., 2005; Whitney, Kirk, O'Sullivan, Ralph, & Jefferies, 2011), some argue that these effects instead reflect attempts to detect and resolve semantic incongruencies or ambiguities (Noppeney et al., 2008). Both views are compatible with our data: whether the heightened activity reflects increased effort to integrate disparate messages or reflects error detection and resolution, both processes require an integrated representation of meaningful information from two simultaneously perceived streams.

Decreased activation for speech with gesture

In addition to activation increases for comprehension of speech with gestures relative to unimodal perception, we searched for regions showing the opposite pattern, reasoning that when speech and gesture convey similar or complementary semantic information, each may aid comprehension of the other, leading to recruitment of fewer neural resources. Relative to gesture alone, speech with gesture led to reduced activation in right middle temporal gyrus and parietal cortex. In these regions associated with action representations, the decreased activation suggests that gesture comprehension is less reliant on action simulations when perceived in the context of speech. Stated another way, the meaningful, representational gestures we presented are not easily interpretable when perceived in isolation, which may lead to increased simulations as the perceiver attempts to discern their meanings or goals.

Similarly, relative to speech alone, small regions in bilateral insula showed reduced activation for speech with gestures. Decreased activity here may reflect reduced processing demand for speech comprehension when meaningful gestures accompany speech. In other words, we hypothesize that semantically related information in another modality renders speech comprehension more efficient. Consistent with this interpretation, Kircher et al. (2009) reported reduced left insula activation for abstract sentences with metaphoric gestures, relative to speech alone, focused approximately 1cm inferior to the cluster identified here. They concluded that the reduced activation reflected less effortful processing when metaphoric gestures help disambiguate abstract sentences. Our findings extend those

of Kircher et al. (2009) by demonstrating that disambiguation is not a critical factor for facilitating multimodal semantic processing. Rather, gesture-related reductions in both studies may reflect reduced processing demand for speech comprehension whenever semantically related co-speech gestures are present (e.g. Hostetter, 2011).

The activation decreases we observed for multimodal co-speech gesture echo a recent report from our laboratory that examined neural responses to code-blends, which are a phenomenon unique to hearing users of signed languages (i.e. bimodal bilinguals; Weisberg et al., 2015). Code-blends are the simultaneous production of manual signs and spoken words. The majority of code-blends convey similar semantic information in sign and speech, such as producing the word 'bird' and the sign 'BIRD' simultaneously (Emmorey, Borinstein, Thompson, & Gollan, 2008; Bishop, 2010). Weisberg et al. (2015) found that hearing signers showed reduced activation during code-blend comprehension, relative to unimodal sign perception, in bilateral prefrontal (including left precentral gyrus) and visual extrastriate regions, and relative to unimodal speech perception, in auditory association cortex (superior temporal gyrus). Thus, code-blends and co-speech gesture perception can both lead to greater neural efficiency compared to unimodal speech or gesture/sign perception. In the case of code-blends we proposed a mechanism involving a redundant signals effect: the semantic information was completely redundant across the visual (sign) and auditory (speech) modalities for code-blends. Furthermore, each unimodal input comprising a code-blend is a fully conventionalized linguistic unit. In contrast, representational gestures are not semantically identical to speech and require context from speech for comprehension. Given these divergent characteristics, we speculate that different mechanisms are behind the reductions associated with multimodal effects in each case (e.g., re-allocation of resources for co-speech gestures vs. redundant signals effect for code-blends).

Co-speech gesture perception in native bimodal bilinguals

We included a group of hearing signers in the current study because we hypothesized that participants with and without life-long sign language experience would show differential responses to co-speech gestures, but this was not the case. Manual movements are common to co-speech gesture and sign language, and code-blends are a natural and frequent phenomenon in bimodal bilingual communication (Baker & Van den Bogaerde, 2008; Bishop, 2010; Petitto et al., 2001). Thus, in addition to typical co-speech gesture perception, bimodal bilinguals have unique experience integrating manual signs with speech. Previous work demonstrating that sign language experience affects co-speech gesture production (Casey & Emmorey, 2009; Casey et al., 2012) hinted that gestures might carry particular salience for hearing signers, which could impact brain systems recruited for multimodal semantic integration. However, this hypothesis was not supported, indicating that the frequent perception of "co-speech signs" does not impact the neural systems engaged in processing co-speech gestures. We speculate that early and concurrent acquisition of speech and sign may lead to the development of relatively segregated systems for processing sign language and co-speech gesture. However, the sequential acquisition of a spoken and then a signed language (as a second language) might lead to less segregated, more interactive systems (Casey et al., 2012). In that case, the interaction of these systems may modulate

brain responses in late signers, compared to non-signers, during co-speech gesture perception. However, this interpretation must remain tentative, as it is possible that our direct comparison of bimodal bilinguals and monolinguals lacked sufficient statistical power to detect subtle group differences. While we cannot rule out this possibility, it nevertheless appears that native acquisition of a visuomanual language does not robustly alter the neural signature of representational co-speech gesture perception.

In summary, we identified a bilateral set of regions incorporating posterior superior temporal gyrus, precentral gyrus, and inferior frontal gyrus that supports the integration of speech with spontaneous, meaningful co-speech gestures. Within the left superior temporal and premotor regions, we identified functionally distinct loci with superadditive responses. We hypothesize that these focal sectors within a larger network play a dual role, supporting both sensory and semantic integration, while the larger, surrounding regions with subadditive responses support integration at either the sensory or the semantic level, but not both. Further studies with higher spatial resolution and paradigms that isolate sensory from semantic integration will shed further light on the circumstances under which, and processes for which each region is recruited. In other regions, reduced activation for audiovisual perception of speech with gesture, relative to unimodal inputs, suggests greater neural efficiency when speech and gesture are perceived together rather than alone, likely due to complimentary information reducing comprehension effort.

Our stimuli have a degree of ecological validity that is unique among co-speech gesture studies — most prior studies presented scripted, manipulated, and rehearsed stimuli with nonsense or incongruent gestures as control conditions. Although such studies afford experimental control to test very specific hypotheses about different gesture types, unnatural contexts may impact neural activity and interpretation. We avoided these challenges by presenting natural, spontaneous co-speech gestures, yet our results largely agree with published studies, supporting the use of strategically controlled stimuli in future investigations. However, the superadditive responses we identified in posterior temporal and premotor cortices may be an effect more readily observed for more ecologically valid stimuli. Our findings indicate that paradigms using spontaneously produced, real world stimuli may detect subtle integration-related activity more readily than those using scripted stimuli, and we encourage more work along those lines.

We have confirmed previous reports that integration of audiovisually presented manual gestures and auditory speech relies on a set of regions that includes the posterior superior temporal, premotor, and inferior frontal cortices. We extended those reports by identifying superadditive responses in circumscribed regions within the posterior superior temporal and premotor areas. These discrete regions may serve a unique functional role in co-speech gesture comprehension, integrating information across both the sensory domain and the semantic domain to enable efficient comprehension of speech accompanied by representational gestures.

To conclude, representational co-speech gestures enhance communication by adding semantic nuance to information conveyed by speech. Integration across the senses and across semantic messages received via different senses enable unified representations to

emerge, enhancing memory, attention, communication, and social interaction. As we gain a deeper understanding of how the brain accomplishes integration across sensory and cognitive domains, an expanding focus of research may include information from additional senses (e.g. tactile, olfactory) and cognitive processes (e.g. memory encoding). Our findings emphasize the need to include multiple and varied paradigms to arrive at a unified account of the neural network supporting integration, accounting for findings from both ecologically valid stimuli and carefully controlled experimental manipulations.

Acknowledgments

NIH Grant HD047736 to SDSU and Karen Emmorey supported this research. We thank Cindy O'Grady Farnady for help recruiting participants, Stephen McCullough for help with data collection, and all who participated in the study, without whom this research would not be possible.

References

- Andric M, Small SL. Gesture's neural language. *Frontiers in Psychology*. 2012; 3doi: 10.3389/fpsyg.2012.00099
- Baker, A., Van den Bogaerde, B. Sign bilingualism: Language development, interaction, and maintenance in sign language contact situations. Plaza-Pust, C., Morales Lopéz, E., editors. Amsterdam: John Benjamins; 2008.
- Barracough NE, Xiao DK, Baker CI, Oram MW, Perrett DI. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience*. 2005; 17(3):377–391. DOI: 10.1162/0898929053279586 [PubMed: 15813999]
- Barsalou LW. Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B-Biological Sciences*. 2009; 364(1521):1281–1289. DOI: 10.1098/rstb.2008.0319
- Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*. 2004a; 41(5):809–823. DOI: 10.1016/s0896-6273(04)00070-4 [PubMed: 15003179]
- Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. *Nature Neuroscience*. 2004b; 7(11): 1190–1192. DOI: 10.1038/nn1333 [PubMed: 15475952]
- Beauchamp MS. Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics*. 2005; 3(2):93–113. DOI: 10.1385/ni:03:02:93 [PubMed: 15988040]
- Bishop M. Happen can't hear: An analysis of code-blends in hearing, native signers of American Sign Language. *Sign Language Studies*. 2010; 11(2):205–240. DOI: 10.1353/sls.2010.0007
- Bookheimer, S. Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. In: Cowan, WM, Hyman, SE, Jessell, TM., Stevens, CF., editors. *Annual Review of Neuroscience*. Vol. 25. 2002. p. 151-188.
- Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Freund HJ. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*. 2001; 13(2):400–404. DOI: 10.1046/j.1460-9568.2001.01385.x [PubMed: 11168545]
- Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E. Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. *Cognitive Brain Research*. 2001; 10(3):349–353. DOI: 10.1016/s0926-6410(00)00054-9 [PubMed: 11167060]
- Callan DE, Jones JA, Munhall K, Kroos C, Callan AM, Vatikiotis-Bateson E. Multisensory integration sites identified by perception of spatial wavelet filtered visual speech gesture information. *Journal of Cognitive Neuroscience*. 2004; 16(5):805–816. DOI: 10.1162/089892904970771 [PubMed: 15200708]

- Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*. 2000; 10(11):649–657. DOI: 10.1016/s0960-9822(00)00513-3 [PubMed: 10837246]
- Calvert GA. Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*. 2001; 11(12):1110–1123. DOI: 10.1093/cercor/11.12.1110 [PubMed: 11709482]
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage*. 2001; 14(2):427–438. DOI: 10.1006/nimg.2001.0812 [PubMed: 11467916]
- Campbell R, MacSweeney M, Surguladze S, Calvert G, McGuire P, Suckling J, David AS. Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). *Cognitive Brain Research*. 2001; 12(2):233–243. DOI: 10.1016/s0926-6410(01)00054-4 [PubMed: 11587893]
- Casey S, Emmorey K. Co-speech gesture in bimodal bilinguals. *Language and Cognitive Processes*. 2009; 24(2):290–312. DOI: 10.1080/01690960801916188
- Casey S, Emmorey K, Larrabee H. The effects of learning American Sign Language on co-speech gesture. *Bilingualism-Language and Cognition*. 2012; 15(4):677–686. DOI: 10.1017/s1366728911000575
- Chen G, Saad ZS, Nath AR, Beauchamp MS, Cox RW. FMRI group analysis combining effect estimates and their variances. *Neuroimage*. 2012; 60(1):747–765. DOI: 10.1016/j.neuroimage.2011.12.060 [PubMed: 22245637]
- Dahl CD, Logothetis NK, Kayser C. Spatial Organization of Multisensory Responses in Temporal Association Cortex. *Journal of Neuroscience*. 2009; 29(38):11924–11932. DOI: 10.1523/jneurosci.3437-09.2009 [PubMed: 19776278]
- Dick AS, Goldin-Meadow S, Hasson U, Skipper JI, Small SL. Co-Speech Gestures Influence Neural Activity in Brain Regions Associated With Processing Semantic Information. *Human Brain Mapping*. 2009; 30(11):3509–3526. DOI: 10.1002/hbm.20774 [PubMed: 19384890]
- Dick AS, Goldin-Meadow S, Solodkin A, Small SL. Gesture in the developing brain. *Developmental Science*. 2012; 15(2):165–180. DOI: 10.1111/j.1467-7687.2011.01100.x [PubMed: 22356173]
- Emmorey K, Borinstein HB, Thompson R, Gollan TH. Bimodal bilingualism. *Bilingualism: Language and Cognition*. 2008; 11(1):43–61. DOI: 10.1017/S1366728907003203
- Emmorey K, McCullough S, Mehta S, Grabowski TJ. How sensory-motor systems impact the neural organization for language: direct contrasts between spoken and signed language. *Frontiers in Psychology*. 2014; 5:13.doi: 10.3389/fpsyg.2014.00484 [PubMed: 24478744]
- Goodwin C. Gestures as a resource for the organization of mutual orientation. *Semiotica*. 1986; 62(1–2):29–49. DOI: 10.1515/semi.1986.62.1-2.29
- Gottfried JA, Dolan RJ. The nose smells what the eye sees: Crossmodal visual facilitation of human olfactory perception. *Neuron*. 2003; 39(2):375–386. DOI: 10.1016/s0896-6273(03)00392-1 [PubMed: 12873392]
- Grafton ST, Fadiga L, Arbib MA, Rizzolatti G. Premotor cortex activation during observation and naming of familiar tools. *Neuroimage*. 1997; 6(4):231–236. DOI: 10.1006/nimg.1997.0293 [PubMed: 9417966]
- Green A, Straube B, Weis S, Jansen A, Willmes K, Konrad K, Kircher T. Neural integration of iconic and unrelated coverbal gestures: A functional MRI study. *Human Brain Mapping*. 2009; 30(10):3309–3324. DOI: 10.1002/hbm.20753 [PubMed: 19350562]
- Hauk O, Johnsrude I, Pulvermüller F. Somatotopic representation of action words in human motor and premotor cortex. *Neuron*. 2004; 41(2):301–307. DOI: 10.1016/s0896-6273(03)00838-9 [PubMed: 14741110]
- Hickok G, Houde J, Rong F. Sensorimotor Integration in Speech Processing: Computational Basis and Neural Organization. *Neuron*. 2011; 69(3):407–422. DOI: 10.1016/j.neuron.2011.01.019 [PubMed: 21315253]
- Hickok G, Poeppel D. Opinion - The cortical organization of speech processing. *Nature Reviews Neuroscience*. 2007; 8(5):393–402. DOI: 10.1038/nrn2113 [PubMed: 17431404]

- Holle H, Gunter TC, Rueschemeyer S-A, Hennenlotter A, Iacoboni M. Neural correlates of the processing of co-speech gestures. *Neuroimage*. 2008; 39(4):2010–2024. DOI: 10.1016/j.neuroimage.2007.10.055 [PubMed: 18093845]
- Holle H, Obleser J, Rueschemeyer S-A, Gunter TC. Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage*. 2010; 49(1):875–884. DOI: 10.1016/j.neuroimage.2009.08.058 [PubMed: 19733670]
- Hostetter AB. When do gestures communicate? A meta-analysis. *Psychological Bulletin*. 2011; 137(2): 297–315. DOI: 10.1037/a0022128 [PubMed: 21355631]
- Hubbard AL, McNealy K, Scott-Van Zeeland AA, Callan DE, Bookheimer SY, Dapretto M. Altered integration of speech and gesture in children with autism spectrum disorders. *Brain and Behavior*. 2012; 2(5):606–619. DOI: 10.1002/brb3.81 [PubMed: 23139906]
- Hubbard AL, Wilson SM, Callan DE, Dapretto M. Giving speech a hand: Gesture modulates activity in auditory cortex during speech perception. *Human Brain Mapping*. 2009; 30(3):1028–1037. DOI: 10.1002/hbm.20565 [PubMed: 18412134]
- Jeannerod M. Neural simulation of action: A unifying mechanism for motor cognition. *Neuroimage*. 2001; 14(1):S103–S109. DOI: 10.1006/nimg.2001.0832 [PubMed: 11373140]
- Josse G, Joseph S, Bertasi E, Giraud A-L. The brain's dorsal route for speech represents word meaning: Evidence from gesture. *Plos One*. 2012; 7(9)doi: 10.1371/journal.pone.0046108
- Kelly SD, Barr DJ, Church RB, Lynch K. Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*. 1999; 40(4):577–592. DOI: 10.1006/jmla.1999.2634
- Kendon A. Gesture. *Annual Review of Anthropology*. 1997; 26:109–128. DOI: 10.1146/annurev.anthro.26.1.109
- Kircher T, Straube B, Leube D, Weis S, Sachs O, Willmes K, Green A. Neural interaction of speech and gesture: Differential activations of metaphoric co-verbal gestures. *Neuropsychologia*. 2009; 47(1):169–179. DOI: 10.1016/j.neuropsychologia.2008.08.009 [PubMed: 18771673]
- Kohler E, Keysers C, Umiltà MA, Fogassi L, Gallese V, Rizzolatti G. Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*. 2002; 297(5582):846–848. DOI: 10.1126/science.1070311 [PubMed: 12161656]
- Laird AR, Robinson JL, McMillan KM, Tordesillas-Gutierrez D, Moran ST, Gonzales SM, Lancaster JL. Comparison of the disparity between Talairach and MNI coordinates in functional neuroimaging data: Validation of the Lancaster transform. *Neuroimage*. 2010; 51(2):677–683. DOI: 10.1016/j.neuroimage.2010.02.048 [PubMed: 20197097]
- Lancaster JL, Tordesillas-Gutierrez D, Martinez M, Salinas F, Evans A, Zilles K, Fox PT. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Human Brain Mapping*. 2007; 28(11):1194–1205. DOI: 10.1002/hbm.20345 [PubMed: 17266101]
- Laurienti PJ, Perrault TJ, Stanford TR, Wallace MT, Stein BE. On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research*. 2005; 166(3–4):289–297. DOI: 10.1007/s00221-005-2370-2 [PubMed: 15988597]
- MacSweeney M, Woll B, Campbell R, Calvert GA, McGuire PK, David AS, Brammer MJ. Neural correlates of British sign language comprehension: Spatial processing demands of topographic language. *Journal of Cognitive Neuroscience*. 2002; 14(7):1064–1075. DOI: 10.1162/089892902320474517 [PubMed: 12419129]
- Marstaller L, Burianova H. The multisensory perception of co-speech gestures - A review and meta-analysis of neuroimaging studies. *Journal of Neurolinguistics*. 2014; 30:69–77. DOI: 10.1016/j.jneuroling.2014.04.003
- Martin A. The representation of object concepts in the brain. *Annual Review of Psychology*. 2007; 58:25–45. DOI: 10.1146/annurev.psych.57.102904.190143
- Martin, A. Circuits in mind: The neural foundations for object concepts. In: Gazzaniga, M., editor. *The cognitive neurosciences*. 4. Boston: MIT Press; 2009. p. 1031–1045.
- Marstaller L, Burianova H. The multisensory perception of co-speech gestures - A review and meta-analysis of neuroimaging studies. *Journal of Neurolinguistics*. 2014; 30:69–77. DOI: 10.1016/j.jneuroling.2014.04.003

- McNeill, D. Hand and mind: What gestures reveal about thoughts. Chicago, IL: University of Chicago; 1992.
- Meister IG, Buelte D, Staedtgen M, Boroojerdi B, Sparing R. The dorsal premotor cortex orchestrates concurrent speech and fingertapping movements. *European Journal of Neuroscience*. 2009; 29(10): 2074–2082. DOI: 10.1111/j.1460-9568.2009.06729.x [PubMed: 19453637]
- Meredith MA, Stein BE. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*. 1986; 56(3):640–662. [PubMed: 3537225]
- Morsella E, Krauss RM. The role of gestures in spatial working memory and speech. *American Journal of Psychology*. 2004; 117(3):411–424. DOI: 10.2307/4149008 [PubMed: 15457809]
- Moss HE, Abdallah S, Fletcher P, Bright P, Pilgrim L, Acres K, Tyler LK. Selecting among competing alternatives: Selection and retrieval in the left inferior frontal gyrus. *Cerebral Cortex*. 2005; 15(11): 1723–1735. DOI: 10.1093/cercor/bhi049 [PubMed: 15728742]
- Murata A, Fadiga L, Fogassi L, Gallese V, Raos V, Rizzolatti G. Object representation in the ventral premotor cortex (area F5) of the monkey. *Journal of Neurophysiology*. 1997; 78(4):2226–2230. [PubMed: 9325390]
- Noppeney U, Josephs O, Hocking J, Price CJ, Friston KJ. The effect of prior visual information on recognition of speech and sounds. *Cerebral Cortex*. 2008; 18(3):598–609. DOI: 10.1093/cercor/bhm091 [PubMed: 17617658]
- Ozyurek A. Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. 2014; 369(1651):20130296–20130296. DOI: 10.1098/rstb.2013.0296 [PubMed: 25092664]
- Petitto LA, Katerelos M, Levy BG, Gauna K, Tetreault K, Ferraro V. Bilingual signed and spoken language acquisition from birth: implications for the mechanisms underlying early bilingual language acquisition. *Journal of Child Language*. 2001; 28(2):453–496. [PubMed: 11449947]
- Pulvermuller F, Fadiga L. Active perception: sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*. 2010; 11(5):351–360. DOI: 10.1038/nrn2811 [PubMed: 20383203]
- Raij T, Uutela K, Hari R. Audiovisual integration of letters in the human brain. *Neuron*. 2000; 28(2): 617–625. DOI: 10.1016/s0896-6273(00)00138-0 [PubMed: 11144369]
- Rizzolatti G, Camarda R, Fogassi L, Gentilucci M, Luppino G, Matelli M. Functional organization of inferior area 6 in the macaque monkey II. Area F5 and the control of distal movements. *Experimental Brain Research*. 1988; 71(3):491–507. DOI: 10.1007/bf00248742 [PubMed: 3416965]
- Scott SK, Johnsrude IS. The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*. 2003; 26(2):100–107. DOI: 10.1016/s0166-2236(02)00037-1 [PubMed: 12536133]
- Seltzer B, Pandya DN. Parietal, temporal, and occipital projections to cortex of the superior temporal sulcus in the rhesus monkey: A retrograde tracer study. *Journal of Comparative Neurology*. 1994; 343(3):445–463. DOI: 10.1002/cne.903430308 [PubMed: 8027452]
- Skipper JI, Goldin-Meadow S, Nusbaum HC, Small SL. Gestures orchestrate brain networks for language understanding. *Current Biology*. 2009; 19(8):661–667. DOI: 10.1016/j.cub.2009.02.051 [PubMed: 19327997]
- Skipper JI, Nusbaum HC, Small SL. Listening to talking faces: motor cortical activation during speech perception. *Neuroimage*. 2005; 25(1):76–89. DOI: 10.1016/j.neuroimage.2004.11.006 [PubMed: 15734345]
- Söderfeldt B, Ingvar M, Rönnerberg J, Eriksson L, Serrander M, Stone-Elander S. Signed and spoken language perception studied by positron emission tomography. *Neurology*. 1997; 49(1):82–87. [PubMed: 9222174]
- Stein BE, Meredith MA, Stein BE, Meredith MA. The merging of the senses. 1993
- Stevenson RA, Geoghegan ML, James TW. Superadditive BOLD activation in superior temporal sulcus with threshold non-speech objects. *Experimental Brain Research*. 2007; 179(1):85–95. DOI: 10.1007/s00221-006-0770-6 [PubMed: 17109108]

- Straube B, Green A, Bromberger B, Kircher T. The differentiation of iconic and metaphoric gestures: Common and unique integration processes. *Human Brain Mapping*. 2011; 32(4):520–533. DOI: 10.1002/hbm.21041 [PubMed: 21391245]
- Straube B, Green A, Weis S, Kircher T. A supramodal neural network for speech and gesture semantics: An fMRI study. *Plos One*. 2012; 7(11)doi: 10.1371/journal.pone.0051207
- Straube B, He Y, Steines M, Gebhardt H, Kircher T, Sammer G, Nagels A. Supramodal neural processing of abstract information conveyed by speech and gesture. *Frontiers in Behavioral Neuroscience*. 2013; 7(120):1–14. DOI: 10.3389/fnbeh.2013.00120 [PubMed: 23423702]
- Szyck GR, Stadler J, Tempelmann C, Muentel TF. Examining the McGurk illusion using high-field 7 Tesla functional MRI. *Frontiers in Human Neuroscience*. 2012; 6doi: 10.3389/fnhum.2012.00095
- Talairach, J., Tournoux, P. Co-planar stereotaxic atlas of the human brain. New York: Thieme; 1988.
- Thompson-Schill SL. Neuroimaging studies of semantic memory: inferring "how" from "where". *Neuropsychologia*. 2003; 41(3):280–292. DOI: 10.1016/s0028-3932(02)00161-6 [PubMed: 12457754]
- Thompson-Schill SL, D'Esposito M, Aguirre GK, Farah MJ. Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences of the United States of America*. 1997; 94(26):14792–14797. DOI: 10.1073/pnas.94.26.14792 [PubMed: 9405692]
- Tourville JA, Reilly KJ, Guenther FH. Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*. 2008; 39(3):1429–1443. DOI: 10.1016/j.neuroimage.2007.09.054 [PubMed: 18035557]
- van Atteveldt N, Formisano E, Goebel R, Blomert L. Integration of letters and speech sounds in the human brain. *Neuron*. 2004; 43(2):271–282. DOI: 10.1016/j.neuron.2004.06.025 [PubMed: 15260962]
- van Atteveldt NM, Blau VC, Blomert L, Goebel R. fMR-adaptation indicates selectivity to audiovisual content congruency in distributed clusters in human superior temporal cortex. *BMC Neuroscience*. 2010; 11doi: 10.1186/1471-2202-11-11
- Weisberg J, McCullough S, Emmorey K. Simultaneous perception of a spoken and a signed language: The brain basis of ASL-English code-blends. *Brain and Language*. 2015; 147:96–106. DOI: 10.1016/j.bandl.2015.05.006 [PubMed: 26177161]
- Weisberg J, van Turenout M, Martin A. A neural system for learning about object function. *Cerebral Cortex*. 2007; 17(3):513–521. DOI: 10.1093/cercor/bhj176 [PubMed: 16581980]
- Werner S, Noppeney U. Superadditive Responses in Superior Temporal Sulcus Predict Audiovisual Benefits in Object Categorization. *Cerebral Cortex*. 2010; 20(8):1829–1842. DOI: 10.1093/cercor/bhp248 [PubMed: 19923200]
- Whitney C, Kirk M, O'Sullivan J, Ralph MAL, Jefferies E. The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. *Cerebral Cortex*. 2011; 21(5):1066–1075. DOI: 10.1093/cercor/bhq180 [PubMed: 20851853]
- Willems RM, Hagoort P, Casasanto D. Body-specific representations of action verbs: Neural evidence from right- and left-handers. *Psychological Science*. 2010; 21(1):67–74. DOI: 10.1177/0956797609354072 [PubMed: 20424025]
- Willems RM, Ozyurek A, Hagoort P. When language meets action: The neural integration of gesture and speech. *Cerebral Cortex*. 2007; 17(10):2322–2333. DOI: 10.1093/cercor/bhl141 [PubMed: 17159232]
- Willems RM, Ozyurek A, Hagoort P. Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage*. 2009; 47(4):1992–2004. DOI: 10.1016/j.neuroimage.2009.05.066 [PubMed: 19497376]
- Witteman J, van Ijendoorn MH, van de Velde D, van Heuven V, Schiller NO. The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia*. 2011; 49(13):3722–3738. DOI: 10.1016/j.neuropsychologia.2011.09.028 [PubMed: 21964199]

Yang J, Andric M, Mathew MM. The neural basis of hand gesture comprehension: A meta-analysis of functional magnetic resonance imaging studies. *Neuroscience and Biobehavioral Reviews*. 2015; 57:88–104. DOI: 10.1016/j.neubiorev.2015.08.006 [PubMed: 26271719]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.
Schematic of audiovisual stimulus presentation paradigm.

Figure 2.

A) Group averaged activation map ($q < 1.0^{-5}$) shows bilateral clusters along the STG, where responses for Speech > Gesture (orange-yellow regions), and in bilateral parietal and occipital cortices, where responses for Gesture > Speech (blue colored regions). B) Group averaged conjunction map displays regions active during perception of Speech alone and during perception of Gesture alone (green), each contrasted with the baseline condition ($q < 1.0^{-5}$). Not visible are overlapping activations in bilateral occipital cortex. Activations displayed on sagittal slices of a template brain.

Figure 3.

Super- and subadditive responses for Speech with Gesture. Superadditive responses were focused in left premotor and superior temporal cortices (blue colored regions; Speech with meaningful Gestures $>$ sum (Speech alone + Gesture alone). Subadditive responses occurred in bilateral occipital cortices (yellow/orange regions; Speech with meaningful Gestures $<$ sum (Speech alone + Gesture alone). Group averaged activations ($q < .05$) are depicted on slices of a template brain.

Table 1

Significantly active clusters for the contrast of Speech vs. Gesture ($q < 10^{-5}$), and their conjunction ($q < .05$). Shown are the nearest Brodmann areas (BA), cluster volumes, Talairach coordinates for center of mass, and peak t-values for each cluster.

Region	BA	Volume (mm ³)	x	y	z	t-value
Speech > Gesture						
L STG	BA 22	29026	-50	-15	3	13.86
R STG	BA 22	22381	52	-16	3	16.46
Gesture > Speech						
L middle/inferior occipital gyrus	BA 37	24439	-36	72	-4	11.73
R middle/inferior occipital gyrus	BA 19	20151	35	76	-4	11.62
L superior parietal lobule	BA 40	10076	-38	-49	46	9.24
R superior parietal lobule	BA 40	8918	36	-47	45	9.65
Conjunction						
R occipital pole	BA 18	13591	26	-85	-2	-
L occipital pole	BA 18	12863	-20	-87	-4	-
L IFG (triangularis)	BA 45	5831	-47	28	4	-
L STS	BA 38	4845	-47	-52	11	-
R STS	BA 41	3301	49	-37	14	-
R IFG (triangularis)	BA 45	3044	53	25	6	-
L precentral gyrus	BA 6	2187	-46	-6	47	-
R precentral gyrus	BA 4/6	1458	50	-7	42	-
L/R medial orbital frontal gyrus	BA 11	1415	1	39	-11	-
L supramarginal gyrus	BA 13	986	-48	-37	24	-
L/R superior frontal gyrus	BA 8	729	-6	42	46	-

Table 2

Significantly active clusters ($q < .05$) for the simple contrasts of Speech vs. Speech with Gesture (in speech-responsive regions) and Gesture vs. Speech with Gesture (in gesture-responsive regions). Shown are nearest Brodmann areas (BA), cluster volumes, Talairach coordinates for center of mass, and peak t-values for each cluster.

Region	BA	Volume (mm ³)	x	y	z	t (max)
Speech > Speech with Gesture						
R insula	BA 13	2144	44	-10	8	3.43
L insula	BA 13	2144	-41	-12	7	4.53
Speech with Gesture > Speech						
L STS	BA 39	22381	-49	-43	13	8.58
L IFG (pars triangularis)	BA 45	14106	-46	21	0	6.21
L lateral occipital gyrus	BA 17	13420	-22	-86	-4	12.17
R lateral occipital gyrus		13206	29	-84	-3	12.15
R STG/planum temporale/supramarginal gyrus	BA 13/42	12905	50	-33	10	6.38
R IFG (pars triangularis)	BA 45	7803	51	20	-3	7.21
L SFG	BA 6	5145	-4	45	42	4.14
L precentral gyrus	BA 6	3344	-46	-7	48	10.31
R precentral gyrus	BA 6	2315	51	-8	43	7.57
L/R superior frontal gyrus	BA 6	1844	-1	-3	62	6.41
R medial orbital/gyrus rectus		815	1	41	-9	3.37
R temporal pole		686	41	22	-31	3.73
Gesture > Speech with Gesture						
R MTG	BA 37	2487	47	-51	-4	4.37
R superior parietal lobule	BA 5/7	557	6	-61	61	3.57
Speech with Gesture > Gesture						
L STS	BA 41	5917	-47	-52	11	9.31
L IFG (pars triangularis)	BA 45	5016	-48	27	3	7.70
R STS	BA 22	3859	48	-38	13	8.90
R IFG (pars triangularis)	BA 45	2358	54	24	6	5.09
L precentral gyrus	BA 6	2358	-47	-6	47	7.40
R precentral gyrus	BA 6	1672	50	-6	43	5.38

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Region	BA	Volume (mm ³)	x	y	z	t (max)
L fusiform gyrus	BA 36/20	943	-39	-37	-16	3.73
L supramarginal gyrus	BA 40	815	-49	-37	23	4.73
L/R medial orbitofrontal gyrus/gyrus rectus	BA 11/32	643	0	38	-10	4.49
R lateral occipital gyrus	BA 17	557	13	-89	2	3.42
L superior frontal gyrus	BA 8	515	-8	42	46	5.29
L lingual gyrus	BA 18	429	-6	-83	-8	3.60

Table 3

Significant clusters from the contrast of Speech with Gesture vs. sum (Speech alone + Gesture alone). Regions where Speech with Gesture > sum (Speech alone + Gesture alone) revealed superadditive responses (all $q < .05$). Shown are nearest Brodmann areas (BA), cluster volumes, Talairach coordinates for center of mass, and peak t-values for each cluster.

Region	BA	Volume (mm ³)	x	y	z	t-value
Sum (speech + gesture) > Speech with Gesture						
L lateral occipital gyrus	BA 17	4502	-19	-89	-3	5.60
R lateral occipital gyrus	BA 17	1844	20	-93	8	5.43
Speech with Gesture > Sum (speech + gesture)						
L STG/S	BA 13	2916	-53	-54	10	4.87
L precentral gyrus	BA 6	772	-45	-7	48	4.47