Behavioral/Cognitive

Transcranial Magnetic Stimulation over Left Inferior Frontal and Posterior Temporal Cortex Disrupts Gesture-Speech Integration

Wanying Zhao, Kevin Riggs, Igor Schindler, and Henning Holle
Department of Psychology, School of Life Sciences, University of Hull, Hull HU6 7RX, United Kingdom

Language and action naturally occur together in the form of cospeech gestures, and there is now convincing evidence that listeners display a strong tendency to integrate semantic information from both domains during comprehension. A contentious question, however, has been which brain areas are causally involved in this integration process. In previous neuroimaging studies, left inferior frontal gyrus (IFG) and posterior middle temporal gyrus (pMTG) have emerged as candidate areas; however, it is currently not clear whether these areas are causally or merely epiphenomenally involved in gesture-speech integration. In the present series of experiments, we directly tested for a potential critical role of IFG and pMTG by observing the effect of disrupting activity in these areas using transcranial magnetic stimulation in a mixed gender sample of healthy human volunteers. The outcome measure was performance on a Stroop-like gesture task (Kelly et al., 2010a), which provides a behavioral index of gesture-speech integration. Our results provide clear evidence that disrupting activity in IFG and pMTG selectively impairs gesture-speech integration, suggesting that both areas are causally involved in the process. These findings are consistent with the idea that these areas play a joint role in gesture-speech integration, with IFG regulating strategic semantic access via top-down signals acting upon temporal storage areas.

Key words: comprehension; gesture; integration; multimodal; speech; TMS

Significance Statement

Previous neuroimaging studies suggest an involvement of inferior frontal gyrus and posterior middle temporal gyrus in gesture-speech integration, but findings have been mixed and due to methodological constraints did not allow inferences of causality. By adopting a virtual lesion approach involving transcranial magnetic stimulation, the present study provides clear evidence that both areas are causally involved in combining semantic information arising from gesture and speech. These findings support the view that, rather than being separate entities, gesture and speech are part of an integrated multimodal language system, with inferior frontal gyrus and posterior middle temporal gyrus serving as critical nodes of the cortical network underpinning this system.

Introduction

Although spoken communication is often considered a purely auditory-vocal process, research in the past decade has provided unequivocal evidence that speech is fundamentally multimodal (Özyürek, 2014). Across all spoken languages, speakers addition-

ally take advantage of the visuomanual modality during communication in the form of hand gestures. For example, speakers may spontaneously use their hands to outline patterns (e.g., when describing the layout of their house) or reenact actions (making a strumming movement while saying "He played the instrument"). There is now convincing evidence that listeners pick up the additional information provided by gestures (Wu and Coulson, 2007; Kelly et al., 2010a; Gunter et al., 2015), although if asked later, they are usually unable to tell whether a particular piece of information originated in the speech or the gesture channel (Alibali et al., 1997). This suggests that, rather than maintaining separate gestural and speech memory traces, listeners combine semantic information arising from the two modalities into a single coherent semantic representation (Özyürek, 2014).
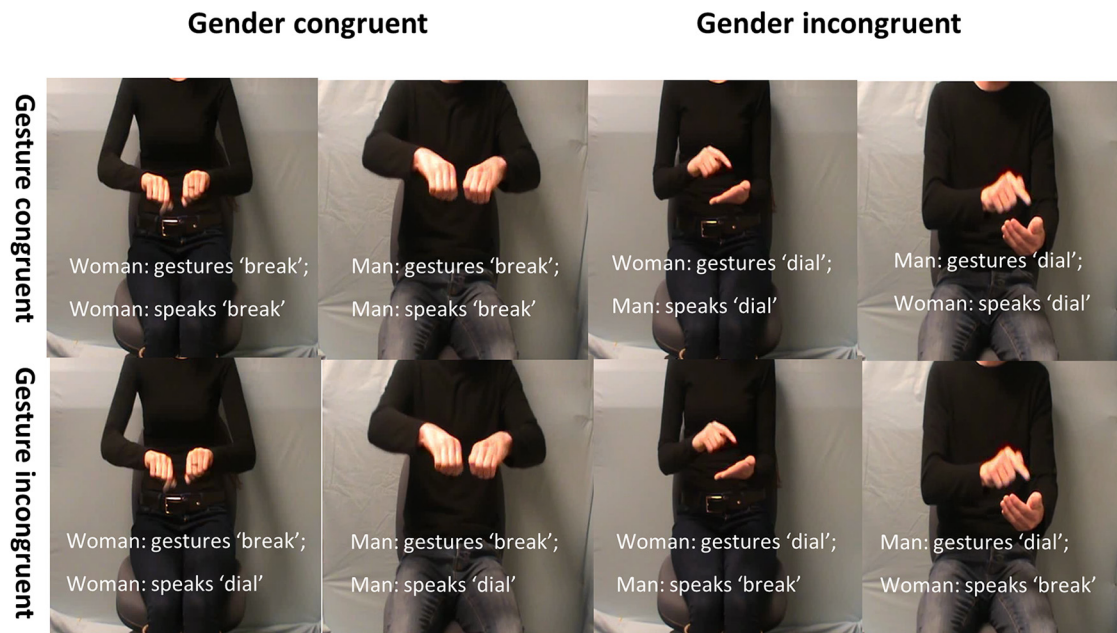
Figure 1. Still frame examples of the experimental video stimuli.

A contentious question over the last decade has been where in the brain the merging of gesture and speech information occurs. Because the information extracted from each modality is qualitatively different (linear, segmented information with arbitrary form-meaning mapping in the case of speech vs holistic, parallel information with motivated form-meaning mapping in the case of gesture), answering this question promises to deepen our understanding of the cortical interface between linguistic and nonlinguistic information.

Previous neuroimaging studies have identified two candidate areas as potential convergence sites: the left inferior frontal gyrus (IFG) and the left posterior middle temporal gyrus (pMTG). To date, however, no consensus has been reached as to which of these areas is causally, and not merely epiphenomenally, involved in this merging process. Some authors have argued for a critical role of the IFG (Willems et al., 2007, 2009), whereas others suggest that pMTG is critically involved (Holle et al., 2008, 2010), with involvement of the IFG restricted to paradigms that induce semantic conflict (Straube et al., 2011). Others have suggested that both areas might be causally involved in linking semantic information extracted from the two domains (Dick et al., 2014). Finally, IFG and pMTG are anatomically well connected (Friederici, 2009), which can produce correlated patterns of activation between these regions (Whitney et al., 2011). It is therefore possible, for example, that activation of pMTG alone is crucial for gesture-speech integration, with IFG activation merely a consequence of its strong anatomical connection with pMTG (or vice versa). fMRI studies are in this respect limited with respect to the degree to which they allow inferences of causality.

In the current study, we used transcranial magnetic stimulation (TMS), a method that is ideally suited to identify causal brain-behavior relationships. This allowed us to disrupt activity in either left IFG or pMTG and observe its effect on gesture-speech integration. The experiment was based on Kelly et al. (2010a), who used the mismatch paradigm to provide a reaction time (RT) index of gesture-speech integration. In their

paradigm, participants were presented with cospeech gestures (e.g., gesturing *typing* while saying "write"), with gender and semantic congruency of audiovisual stimuli being experimentally manipulated. Participants had to identify the gender of the spoken voice. Kelly et al. (2010a) found that, although task irrelevant, gestural information strongly influenced RTs, with participants taking longer to respond when gestures were semantically incongruent with speech. This was interpreted as evidence for the automatic integration of gesture and speech during comprehension.

The aim of the present study was to investigate whether either pMTG and/or IFG are critical for gesture-speech integration. Using the Kelly et al. (2010a) task, we tested whether the magnitude of the semantic congruency effect is reduced when activity in these areas is perturbed using TMS, relative to control site stimulation.

Materials and Methods

Stimuli. Stimuli consisted of 44 simple reenactments of actions (e.g., typing on a keyboard, throwing a ball, zipping up a coat) and were selected based on previous studies on cospeech gestures (Kelly et al., 2010a; Dick et al., 2014; Drijvers and Özyürek, 2017). Each action was produced by either a man or a woman while simultaneously uttering the corresponding speech token (e.g., gesturing *typing* while saying "type"), with only the torso being visible (Fig. 1). Video recordings of these cospeech gestures formed the visual component of our stimulus set. Recordings were subsequently edited so that each video started with the gesture stroke. In a follow-up session, the two volunteers again produced each gesture accompanied by the speech token, but this time only the speech was recorded. Video and audio material were then combined to realize the experimental manipulations of Gender congruency and Semantic congruency. To realize the manipulation of Semantic congruency, a gesture was paired with a seemingly incongruent speech token (e.g., gesturing *ironing* while saying "whisk"). Importantly, the reverse combination was also realized (e.g., gesturing *whisking* while saying "iron"), which ensures that item-specific effects are counterbalanced across the stimulus set.

Pretest 1: semantic congruency rating. To verify that the semantically congruent or incongruent combinations of gesture and speech are indeed perceived as such, a separate set of participants (*n* = 21) rated the

relationship between gesture and speech on a 5 point Likert scale (1 = "no relation"; 5 = "very strong relation"). Based on the rating results, four pairs of stimuli were excluded. The 40 remaining stimuli were used for further pretests. The mean rating for the remaining set of congruent videos was 4.71 (SD 0.32), and the mean rating for the incongruent videos was 1.28 (SD 0.29).

Pretest 2: validation of paradigm and stimulus set. Before the brain stimulation experiments, the RT paradigm was validated in a separate set of participants ($N = 37$), to see whether we could replicate the findings of Kelly et al. (2010a) with our stimulus set. Participants consisted of undergraduate students from the University of Hull, who completed the experiment in exchange for course credit. Participants were asked to indicate, as quickly and as accurately as possible via button press, whether the word in the video was spoken by either a male or a female. Each video started with the onset of a gesture stroke, with the onset of the speech token occurring 200 ms later. Participants made very few errors on the task (overall accuracy >96%); therefore, accuracy data were not statistically analyzed. RTs were calculated relative to the onset of the spoken word. After excluding incorrectly answered trials (3.2%), outliers were determined as those scores that were located in the extreme 5% on either end of the Z-normalized distribution of RTs. This is equivalent to removing scores above and <1.65 SD of a subject's mean RT. Overall, this resulted in 7.3% of trials being excluded as outliers, within the 5%–10% region recommended by Ratcliff (1993). A 2 (Semantic congruency) × 2 (Gender congruency) repeated-measures ANOVA yielded a significant main effect of Gender congruency ($F_{(1,36)} = 63.45$, $p = 1.8563e^{-09}$, Cohen's $d_z = 10.43$), with gender incongruent trials eliciting slower RTs (mean ± SD, 650.68 ± 102.81) than gender congruent trials (625.68 ± 109.50). Crucially, there was also a significant main effect of Semantic congruency ($F_{(1,36)} = 15.12$, $p = 0.0004$, $d_z = 2.49$), indicating that participants were slower to judge the gender of the speaker when speech and gesture were semantically incongruent (643.09 ± 108.34) relative to when they were semantically congruent (632.92 ± 103.84). The interaction of gesture × gender was not significant ($F_{(1,36)} = 1.92$, $p = 0.18$). Thus, we were able to replicate the main finding of Kelly et al. (2010a) in our stimulus set. Participants were slower to judge the gender of the speaker when gesture and speech were semantically incongruent, even though the semantic relationship between gesture and speech was not relevant to the task. The RT cost incurred by semantically incongruent gesture-speech pairs suggests that the representational content of gesture is automatically connected to the representational content of speech.

To maximize the statistical power of the brain stimulation experiments, we also used the results of Pretest 2 to select those item pairs that produced the strongest semantic congruency effect. To this end, we excluded those 4 stimulus pairs from the stimulus set that did not show a semantic congruency effect in the expected direction ($RT_{Sem\_Inc} > RT_{Sem\_Con}$). Thus, the final stimulus set consisted of 32 gestures. Because each gesture was realized either by a male or a female actor, as semantically congruent or incongruent, and also as gender congruent or incongruent, the total stimulus set consisted of 256 videos (32 × 2 × 2 × 2).

Post hoc test: nameability of stimuli. Another independent set of participants ($N = 42$) was asked (after TMS studies had been completed) to provide a verbal label for each gesture, which were presented to them without sound. For each gesture, we then calculated the percentage of participants that provided the correct label. The overall mean nameability index of the final stimulus set of 32 gestures was 49%. This indicates that, as a whole, the stimulus set is best characterized as containing iconic gestures, which are characterized by a certain ambiguity when presented in the absence of speech (Hadar and Pinchas-Zamir, 2004; Drijvers and Özyürek, 2017).

Experiment 1: experimental design and statistical analysis. In Experiment 1, we explored whether disrupting activity in areas hypothesized to underlie gesture-speech integration (left IFG and/or left pMTG) leads to a reduction of the semantic congruency effect. In a within-subject design, participants underwent three sessions, where continuous theta burst

Movie 1. Still 1. Example video 1: Semantically congruent/gender congruent.
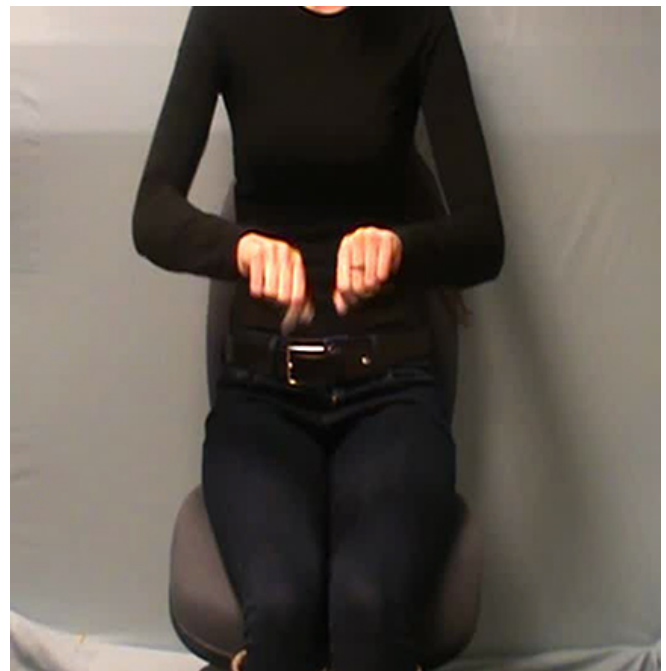
Movie 2. Still 2. Example video 2: Semantically congruent/gender incongruent.

Movie 3. Still 3. Example video 3: Semantically incongruent/gender congruent.

Movie 4. Still 4. Example video 4: Semantically incongruent/gender incongruent.

stimulation (cTBS) was applied to the left IFG, pMTG, or a control site (vertex). The session order was counterbalanced across participants. After stimulation, which occurred at the beginning of each session, participants completed the RT task described above (section Pretest 2). Thus, the full experimental design was a 3 (area: IFG, pMTG, Vertex) × 2 (Gender congruency) × 2 (Semantic congruency) factorial design, and a corresponding 3 × 2 × 2 repeated-measures ANOVA was used to analyze the RT data. Greenhouse–Geisser correction was applied where necessary. For all significant effects, Cohen's d_z (Cohen, 1988) is provided as a standardized effect size measure. We predicted an interaction between Area and Semantic congruency, in the form of a reduction of the semantic congruency effect following either IFG or pMTG stimulation, relative to control site stimulation. The factor of Gender congruency was used as an additional control, to see whether brain stimulation specifically disrupts the processing of semantic (in)congruencies, or more generally interferes with task processing. The size of the critical main effect of semantic congruency, as determined in Pretest 2, was $d_z = 2.49$. An a priori sample size estimation (Faul et al., 2007) indicated that, to detect a TMS-induced reduction of the size of this semantic congruency effect of at least 1 d_z with 95% probability, a sample size of at least 17 participants is required. Accordingly, testing continued until 17 complete datasets were available for analysis.

Experiment 1: participants. Twenty participants took part in Experiment 1 having given written informed consent. Three participants were excluded from the analysis: one for not being able to follow instructions, and another two because of computer malfunction. The experimental protocol was approved by the Ethics Committee of the Department of Psychology. The final sample consisted of 17 participants (6 males and 11 females, age 20–42 years, mean ± SD age, 25.06 ± 5.87 years). All were native English speakers and were classified as right-handed according to the Edinburgh Handedness form (LQ = 73.51, SD = 22.12), had normal or corrected-to-normal vision, and were screened for TMS suitability using a medical questionnaire. Participants received financial compensation at a rate of £8/h.

Experiment 1: stimuli. As mentioned before (see Pretest 2), the final stimulus set consisted of 256 videos, created from 32 gestures (Fig. 1; Movies 1–4). The mean length of a gesture video in the final stimulus set was 1833 ms (SD 259 ms). The mean length of the spoken word was 588 ms (SD 99 ms). Still frame examples of the experimental stimuli are shown in Figure 1.

Experiment 1: procedure. Sessions were scheduled to be at least 1 week apart. In each session, participants were guided to sit in front of a computer and keyboard. After theta-burst stimulation, they were asked to complete the experimental task consisting of 256 gesture videos. Participants received the following instructions:

> "In this experiment, you will observe videos of a person gesturing and speaking at the same time. In some of these videos, the gender of the person you see in the video will be different from the gender of the person you hear (e.g., a male person gestures, but a female voice speaks), whereas in others the gender will be the same. Your task is to indicate, as quickly and accurately as possible, whether the word in a video was spoken by a male or a female."

Participants received a training run of trials to make sure they understood the task. The items used in the training were different to those used in the main experiment. Each trial began with the presentation of a video, with a speech token inserted at 200 ms. For each video, participants had to indicate via button press whether the word in the video was spoken by either a male or a female. If they failed to respond within 2000 ms, the trial was recorded as a time-out and a clock was presented for 500 ms prompting participants to respond faster. In the case of an incorrect answer, a frownie symbol appeared on the screen. Trials were separated by a variable intertrial interval of 500, 1000, or 1500 ms, during which a fixation cross was presented on the screen. The 256 videos of each experimental sessions were presented in blocks of 64 trials each, and participants could take short breaks between blocks. Reponses were made using the left and right
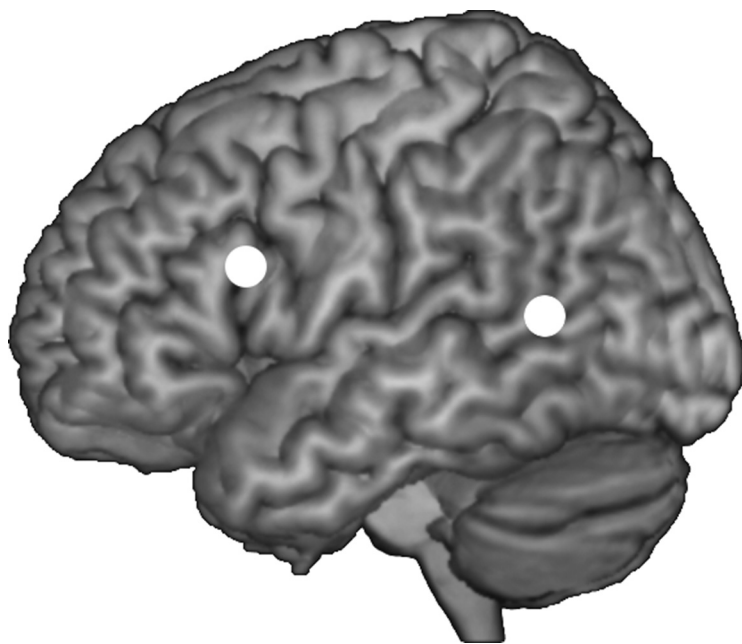
Figure 2. Overview of stimulation sites in MNI space: IFG (−62, 16, 22) and pMTG (−50, −56, 10). Vertex was used as control site.

index finger, and key assignment was counterbalanced across participants, as far as possible. Presentation of videos and collection of RTs was realized using Presentation (www.neurobs.com; RRID:SCR_002521).

Experiment 1: TMS protocol and site localization. The stimulation sites corresponded to MNI coordinates determined in a quantitative meta-analysis using activation likelihood estimation (Eickhoff et al., 2009) of several fMRI studies on iconic-speech integration (Willems et al., 2007, 2009; Holle et al., 2008, 2010; Green et al., 2009; Straube et al., 2011; Dick et al., 2014). Two locations were identified as consistently activated across studies: the left IFG (−62, 16, 22) and the left pMTG (−50, −56, 10).

To enable an image-guided TMS navigation, high-resolution (1 × 1 × 0.6 mm) T1-weighted anatomical MRI scans of each participant were acquired at Hull Royal Infirmary using a GE Medical Systems scanner with a field strength of 3 tesla. MNI coordinates of the target areas were defined as regions of interest (ROIs) using Marsbar (marsbar. sourceforge.net) and SPM12 (Fig. 2). These ROIs were then backprojected from MNI space into each participant's native brain space, using SPM12's inverse transformation function. Subject-specific ROIs were then imported into BrainVoyager (RRID:SCR_013057) and superimposed on the surface reconstruction of the two hemispheres and defined as targets during neuronavigation. This ensured precise stimulation of each target region in each participant.

A Magstim Rapid^2 stimulator was used to generate repetitive magnetic pulses. The pulses were delivered with a standard 70 mm figure-8 coil. A cTBS train of 804 pulses was used (268 bursts, each burst consisting of three pulses at 30 Hz, repeated at intervals of 100 ms, lasting for 40 s) (Nyffeler et al., 2008, 2009). We decided to use a fixed stimulation intensity of 40% of maximum machine output for all participants instead of individual motor-threshold-related intensities because previous studies indicated that the motor threshold is not an appropriate measure for determining stimulation intensity over a nonmotor area (Stewart et al., 2001). The value of 40% was determined in piloting studies as the maximum intensity still tolerated by our participants.

Experiment 2: experimental design and statistical analysis. In Experiment 1, the effect of brain stimulation to pMTG on the semantic congruency effect did not reach full significance ($p = 0.057$, see Results). To further investigate a possible role for pMTG, Experiment 2 was conducted using online TMS, as opposed to offline cTBS, to disrupt brain activity. Online TMS, in the form of repetitive transcra-

nial magnetic stimulation (rTMS), has the advantage that perturbation of cortical activity can be synchronized with the presentation of the experimental stimuli. This enables a more powerful statistical comparison of the effects of brain stimulation on gesture-speech integration.

Experiment 2 used a 2 (Area: pMTG, vertex) × 2 (Gender congruency) × 2 (Semantic congruency) factorial design. We predicted that pMTG would significantly reduce the semantic congruency effect, as indicated by a significant Area × Semantic congruency interaction. Furthermore, we hypothesized that rTMS of pMTG would specifically disrupt gesture-speech integration, but not general task processing, as indicated by an absent interaction of Area and Gender congruency. All other details concerning the statistical analysis were as described for Experiment 1.

Experiment 2: participants. Thirteen participants participated in Experiment 2. One participant was excluded for not following instructions. The final sample used for statistical analysis consisted of 5 males and 7 females (age range: 20–36 years; mean ± SD age, 24.08 ± 4.36 years). All were English native speakers and were classified as right-handed according to the Edinburgh Handedness form questionnaire (Oldfield, 1971), with a Mean Laterality Coefficient of 71.32 (SD 21.42). All other participant details were as described in Experiment 1.

Experiment 2: procedure. Each participant completed four blocks of alternating pMTG or vertex stimulation in a single experimental session. Presentation order was counterbalanced across participants. Within each block, 64 trials were presented. Within each trial, five pulses were delivered at a frequency of 10 Hz for a duration of 500 ms at 45% of maximum machine output. As in Experiment 1, the stimulation intensity was determined in piloting studies as the maximum intensity tolerable by our participants and fixed at 45% for all participants. A stimulation of the left IFG using online 10 Hz rTMS was also initially considered, but not further pursued, because of extreme discomfort and task-distracting effects in form of facial muscle twitches.

The first pulse coincided with the onset of the spoken word. Given the fact that the gesture began 200 ms before speech onset and previous research indicates that semantic gesture-speech integration takes place between 350 and 550 ms after the onset of the gesture stroke (Özyürek et al., 2007), we predicted that such stimulation over a relevant brain region would impair gesture-speech integration. Duration and intensity of the rTMS stimulation were both within the participants' bearable limit and the neuropsychological application safety limit (Anand and Hotson, 2002). All other experimental details were as described in Experiment 1.

Results
Experiment 1
After removing incorrectly answered trials (3.7%) and outliers (6.9%), RT data were subjected to a repeated-measures ANOVA with the factors Area (IFG, pMTG, vertex), Gender congruency (same, different) and Semantic congruency (congruent, incongruent). The ANOVA yielded a significant main effect of Semantic congruency ($F_{(1,16)} = 14.64$, $p = 0.001$, $d_z = 3.55$), reflecting longer RTs or semantically incongruent trials (mean ± SE, 543 ± 16.6) than congruent trials (530 ± 13.7). Furthermore, a significant main effect of Gender congruency ($F_{(1,16)} = 45.37$, $p = 3.49e^{-06}$, $d_z = 11.00$) was observed, indicating that RTs were longer when speech and gesture were produced by different genders (554 ± 15.2) than the same gender (518 ± 15.4). The main effect of Area was not signifi-

Table 1. Reaction times for Experiment 1[a]

	Semantically congruent		Semantically incongruent	
	Gender same	Gender different	Gender same	Gender different
IFG	519 (17)	549 (17)	526 (22)	562 (19)
pMTG	495 (11)	535 (13)	508 (13)	543 (13)
Vertex (control)	521 (17)	559 (19)	540 (20)	578 (22)

[a]Data are mean (SEM) in milliseconds.

cant ($F_{(1.944,30.466)} = 2.965$, $p = 0.065$). Crucially, there was a significant Semantic congruency × Area interaction ($F_{(1.944,30.466)} = 3.53$, $p = 0.042$), indicating that the magnitude of the semantic congruency effect was modulated depending on the brain area stimulated. No such modulation was observed for the gender congruency effect, as indicated by a nonsignificant Area × Gender congruency interaction ($F_{(1.944,30.466)} = 0.50$, $p = 0.60$). The full pattern of results is shown in Table 1.

Simple effects analyses (Fig. 3) indicated that the size of the semantic congruency effect was significantly reduced ($t_{(16)} = 2.58$, $p = 0.020$, $d_z = 1.61$) when cTBS was applied to the left IFG (9.6 ± 4.5), relative to control site stimulation (19.3 ± 5.1). A similar pattern was observed following stimulation of pMTG (10.5 ± 2.3), although this comparison did not reach full significance ($t_{(16)} = 2.05$, $p = 0.057$). There was no evidence that stimulation of either pMTG or IFG modulated the size of the effect of gender congruency (all $t < 0.77$, all $p > 0.451$; Fig. 3).

Experiment 2

After removing incorrectly answered trials (5.8%) and outliers (5.0%), a 2 × 2 × 2 repeated-measures ANOVA of the RT data (Table 2) revealed a significant main effect of Gender congruency ($F_{(1,11)} = 41.36$, $p = 3.25e^{-05}$, $d_z = 11.94$), with longer RTs to gestures and speech from actors of different genders (mean ± SE, 584.7 ± 19.0) than to actors of the same gender (549.6 ± 19.3). The main effect of Area was not significant ($F_{(1,11)} = 1.62$, $p = 0.229$). Crucially, there was a significant interaction of Semantic congruency and Area ($F_{(1,11)} = 9.01$, $p = 0.012$, $d_z = 2.60$), indicating that the magnitude of the semantic congruency effect was modulated by rTMS (Fig. 4). No such effect of brain stimulation on the gender congruency effect was observed, as indicated by a nonsignificant Area × Gender congruency interaction ($F_{(1,11)} = 1.51$, $p = 0.245$). All other main effects or interactions of the ANOVA were not significant (all $F < 3.00$, all $p > 0.111$).

As can be seen in Figure 4, stimulating pMTG in Experiment 2 completely eliminated and actually reversed the semantic congruency effect, relative to control site stimulation. No such effect of brain stimulation was observed for the gender congruency effect, suggesting that rTMS of the pMTG specifically disrupted the semantic integration of gesture and speech.

Discussion

Together, the results of the two studies presented here provide clear evidence that both IFG and pMTG are involved in the merging of semantic information from iconic gestures and speech. When cortical excitability of these areas was decreased via TMS, a reduced RT cost was observed indicating a reduction in semantic integration. TMS of IFG and pMTG was found to specifically disrupt gesture-speech integration, but not general task processing. By directly linking brain activity to behavior, our study demonstrates, for the first time, that both IFG and pMTG are causally involved in integrating information from the two domains of gesture and speech.

Theoretical accounts of cospeech gesture comprehension (McNeill et al., 1994; Kelly et al., 2010b) stress that humans are predisposed to integrate the information from gesture and speech into a single system of meaning. The involuntary and (to some extent) automatic character of this integration process can be inferred from the fact that gesture influences behavioral performance on semantic tasks, even when gestural information is not task relevant (Kelly et al., 2010a,b) or when participants are asked to ignore gesture (Kelly et al., 2007). In the present study, we used a Stroop-like task, where participants were asked to indicate the gender of the spoken voice in each video. Orthogonal to this experimental task, we manipulated semantic integration load, with gesture and speech being either semantically congruent or incongruent. In the unperturbed brain (see Pretest 2 and Vertex condition), this semantic congruency manipulation elicited a RT cost, with longer RTs for semantically incompatible gesture-speech combinations. This finding is in line with the automaticity assumption of gesture-speech integration (McNeill et al., 1994; Kelly et al., 2010b).

Another way to conceptualize the RT costs elicited by the semantic congruency manipulation in the unperturbed brain is in terms of competition for cognitive resources in a limited capacity system (Luber and Lisanby, 2014). The automatic tendency to try to integrate gesture and speech requires additional resources when semantically conflicting stimuli are presented. This results in fewer cognitive resources being available for the orthogonal gender judgment task. In this framework, a disruption of activity in a brain area that is causally contributing to gesture-speech integration should lead to a reduction of the incurred RT cost because the competing/distracting influence of gesture on the experimental task is reduced. This is exactly the pattern we observed in both of the experiments reported here. In Experiment 1, the RT cost attributable to gesture-speech integration was significantly reduced when activity in left IFG was perturbed following offline cTBS. A similar pattern was observed for the left pMTG, although the associated statistical test did not reach full significance ($p = 0.057$). In Experiment 2, we observed that, when activity in pMTG is perturbed using online 10 Hz rTMS, the RT cost related to gesture-speech integration is completely abolished.

The pattern observed in both studies of a reduced RT cost following inhibitory brain stimulation can most readily be explained as a behavioral enhancement via an "addition-by-subtraction" mechanism. As recently summarized by Luber and Lisanby (2014), "addition-by-subtraction" is a mechanism where TMS can produce cognitive enhancement by disrupting processes that compete or distract from task performance. For example, Walsh et al. (1998) observed that, when activity in motion-sensitive area V5 was disrupted, performance was impaired when the task set included judgments of motion direction. However, the same stimulation produced behavioral facilitation when the task set did not require motion judgements (i.e., when participants were only asked to attend to color and form). This suggests a competition between cortical areas, where multiple stimulus attributes are evaluated in parallel (Luber and Lisanby, 2014). When information about movement was not task relevant, disruption of competing but irrelevant movement information (via TMS of V5) decreased total processing time. A similar exemplar of behavioral enhancement via "addition-by-subtraction" is the study by Hayward et al. (2004). They observed that the RT costs in a number Stroop task were reduced when 10 Hz online rTMS was applied to the anterior cingulate cortex.
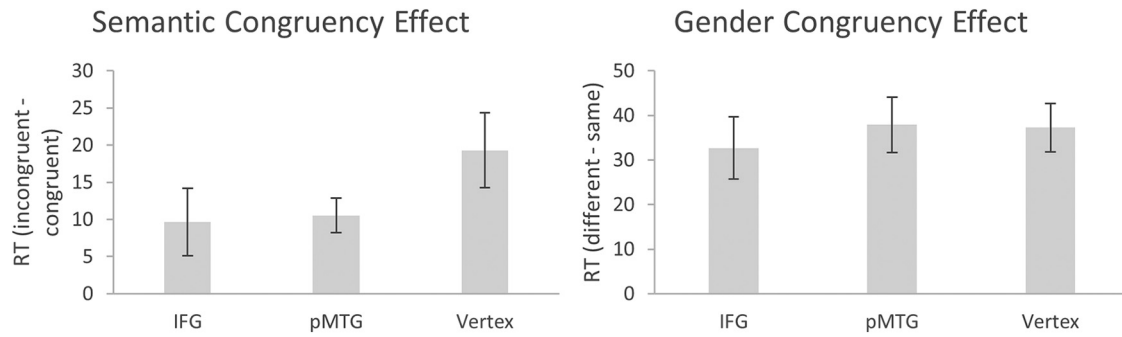
Semantic Congruency Effect

Gender Congruency Effect

Figure 3. Magnitude of semantic and gender congruency effects (ms) for Experiment 1, separately for each stimulation condition. Error bars indicate SEM.

Table 2. Reaction times for Experiment 2[a]

	Semantically congruent		Semantically incongruent	
	Gender same	Gender different	Gender same	Gender different
pMTG	557 (21)	594 (19)	549 (15)	592 (21)
Vertex (control)	537 (25)	568 (17)	556 (19)	584 (23)

[a]Data are mean (SEM) in milliseconds.

Because the conclusions in the present study about the functional role of IFG and pMTG rely on two different TMS protocols (offline cTBS in Experiment 1, online 10 Hz rTMS in Experiment 2), it is important to consider the commonalities and differences in how these different kinds of stimulation affect neural processing and behavior. Both are standard stimulation protocols in TMS research, and each protocol is assumed to perturb normal levels of activity in the stimulated brain region. However, the underlying mechanisms are different. Brief trains of 10 Hz online rTMS are assumed to add random neural noise to the stimulated brain area. Whereas activity in the unperturbed brain area is often organized in synchrony (Allen et al., 2007), rTMS can locally disrupt the phase synchrony underlying this coordinated neural firing, which can have a detrimental influence on behavior (Ruzzoli et al., 2010). The behavioral consequences of online 10 Hz rTMS are usually short-lived and rarely outlast the end of stimulation by more than a few seconds (Luber and Lisanby, 2014). In contrast, cTBS is a protocol where short high-frequency bursts of TMS pulses (e.g., 3 pulses at 50 Hz per burst) are interspersed with brief periods of no stimulation (e.g., one burst every 200 ms). This patterned form of stimulation, delivered for 40 s, has been demonstrated to cause a decrease in cortical excitability lasting up to 60 min (Huang et al., 2005). A recent model suggests that these long-lasting inhibitory aftereffects of cTBS involve LTD-like phenomena (Suppa et al., 2016). One key difference between the two protocols is when the TMS stimulation is applied relative to when RTs are assessed. An advantage of 10 Hz rTMS is that it can be applied online during the time period where the process of interest is taking place. This increases confidence that any TMS effects on behavior are indeed related to the hypothesized mechanism, provided that muscular side effects of TMS are not interfering with task performance (as it was the case when we tried applied online rTMS to IFG, see Experiment 2). Offline cTBS is advantageous in this respect for areas, such as the IFG, because the patterned form of stimulation is better tolerated by participants and the offline nature of the protocol bears the advantage that muscular side effects are less likely to influence behavior on the RT task. A final consideration is whether the different number of trials across the two experiments (Experiment 1: 3 sessions with 256 trials each; Experiment 2: 1 session with 256 trials) affected the

stability of the critical semantic congruency effect over time. However, no significant main effect or interaction involving the factor time was observed in the corresponding ANOVAs, indicating that the semantic congruency effect was not affected and thus did not change over time.

Together, the findings obtained across the two experiments provide clear evidence that both IFG and pMTG are causally contributing to gesture-speech integration during comprehension. Previous brain stimulation studies have already pointed toward the IFG as an important node in the cortical networking mediating the relationship between gesture and speech. For example, Gentilucci et al. (2006) observed that the left IFG is causally involved in linking gesture comprehension and speech production. A very recent paper by Siciliano et al. (2016) demonstrated that the left IFG makes a critical contribution toward a previously demonstrated gesture benefit in foreign language learning. In the present study, we observed that RT costs associated with gesture-speech integration are significantly reduced following perturbation of activity in left IFG using cTBS. As explained below, a potential functional role of the IFG during the comprehension of cospeech iconic gestures could be the strategic recovery of context-appropriate semantic information.

According to Whitney et al. (2011), semantic cognition involves (1) accessing information within the semantic store itself and (2) executive mechanisms that direct semantic activation so that it is appropriate for the current context. In terms of its neural instantiation, it is often assumed that semantic cognition involves modulatory signals from the IFG acting upon temporal storage areas (Lau et al., 2008; Whitney et al., 2011; Yue et al., 2013). Because iconic gestures represent objects and actions by bearing only a partial resemblance to them (Wu and Coulson, 2011), their comprehension may require strategic recovery of semantic activation, to come up with an interpretation of the observed gesture that is compatible with the accompanying speech context. For example, a gesture consisting of two closed fists moving forward from the body center may initially only activate the general concept *push*, but needs additional strategic recovery of semantic activation, via modulatory signals from the IFG acting upon posterior temporal storage areas, to achieve an interpretation that is consistent with the accompanying speech unit *mow*.

Relative to congruent gesture-speech pairs, semantically incongruent combinations trigger an increased need for strategic recovery of semantic information, in an (eventually probably unsuccessful) attempt to resolve the semantic conflict between gesture and speech. Disrupting activity in the IFG interferes with this strategic recovery process, as reflected in the significantly de-
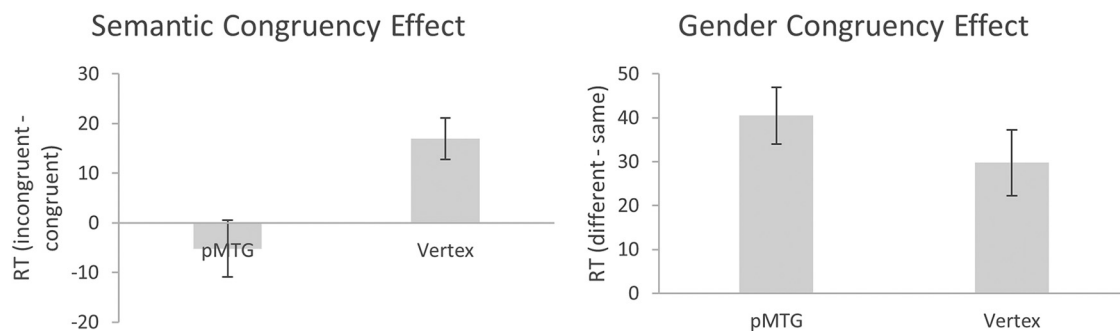
Semantic Congruency Effect

Gender Congruency Effect

Figure 4. Magnitude of semantic and gender congruency effects (ms) for Experiment 2, separately for each stimulation condition. Error bars indicate SEM.

creased RT costs following cTBS of the left IFG. The effect of cTBS on IFG cannot be dismissed as a general disruption of cognitive processing because stimulation of IFG specifically reduced the (task-irrelevant) semantic congruency effect, but not the (task-relevant) gender congruency effect (Fig. 3).

IFG and pMTG most likely work together in integrating gesture with speech, with the above-mentioned modulatory signals originating in the IFG acting upon temporal storage areas. The posterior temporal cortex, encompassing the middle temporal gyrus and adjacent superior temporal sulcus, has been suggested to be involved in accessing semantic information (Lau et al., 2008), either by serving as an interface to a widely distributed network of brain region representing semantic knowledge or by accessing feature knowledge directly stored in pMTG. When presented in isolation, spoken words, lexicalized gestures, as well as less formalized iconic gestures all activate pMTG, which was interpreted that this area is indeed a hub for supramodal access of semantic information (Xu et al., 2009; Straube et al., 2012). Incongruent combinations of gesture and speech most likely place a higher semantic access load on this area than their congruent counterparts. From this perspective, our finding that rTMS of pMTG significantly reduces the size of the semantic congruency effect can be interpreted as an interference in the access of supramodal representations.

Most neurocognitive models of language do not currently consider the influence of extralinguistic semantic information, such as cospeech gestures (Hickok and Poeppel, 2007; Friederici, 2011). The exception is the model by Hagoort (Hagoort and van Berkum, 2007; Hagoort, 2013), which assumes that the integration of gesture and language semantics involves a dynamic interplay between IFG and pMTG, similar to the interpretation of the present findings provided above. The Hagoort model assumes that unification processes, including semantic integration, involve activation of semantic representations in posterior temporal and inferior parietal cortices, as well as modulatory control of the activation level of these representations via a feedback loop between IFG and pMTG. This links well with recent functional connectivity studies (Willems et al., 2009; Yue et al., 2013; Hartwigsen et al., 2017a), which indicate that an increase in semantic integration difficulty leads to an increase in the degree to which IFG exerts control over pMTG. This is the case both for pure linguistic manipulations (e.g., when encountering a semantically anomalous word in a sentence, see Hartwigsen et al., 2017a) and for cospeech gestures (Willems et al., 2009) and may reflect a neural mechanism underpinning inhibition of competing semantic representations (Hartwigsen et al., 2017a). Future studies should include cortical activity and functional connectiv-

ity measures as additional outcome variables because TMS not only affects cortical activity of the perturbed brain area but also changes activity of areas that are functionally connected with the perturbed brain area (Jackson et al., 2016; Hartwigsen et al., 2017b; Wawrzyniak et al., 2017). This will allow insights into the degree to which behavioral changes following perturbation of IFG and pMTG are driven by rapid reorganization of the wider semantic network.

The pMTG and IFG are not only critical for gesture-speech integration, but for semantic cognition more generally. There is considerable evidence demonstrating that these areas are involved in controlled semantic retrieval (e.g., Noonan et al., 2013; Davey et al., 2015), and the present study provides further evidence that this process is independent of modality (Ralph et al., 2017). Interesting in this context is the role of the anterior temporal lobe (ATL). An influential model of semantic cognition, the hub-and-spoke model (Ralph et al., 2017), assumes that this area serves as a transmodal hub for accessing modality-specific conceptual representations distributed across the cortex. However, the ATL shows little or no activation during cospeech gesture comprehension (Marstaller and Burianová, 2014; Özyürek, 2014; Yang et al., 2015), a process that arguably involves combining modality-specific knowledge. More research is needed to clarify whether this lack of ATL activation during gesture-speech integration is indeed a true negative or more merely reflects methodological problems involved in obtaining a good signal-to-noise ratio from the ATL (Visser et al., 2010).

In conclusion, our study provided clear evidence that IFG and pMTG are both critically involved in the integration of gestural and spoken information during comprehension. By linking cortical activity in these areas directly to observed behavior, our study is the first to provide evidence that both areas are causally involved in this process. These findings support the view that, rather than being separate entities, gesture and speech are part of an integrated multimodal language system, with IFG and pMTG serving as critical nodes of the cortical network underpinning this system.

References

Alibali MW, Flevares LM, Goldin-Meadow S (1997) Assessing knowledge conveyed in gesture: do teachers have the upper hand. J Educ Psychol 89:183–193. CrossRef

Allen EA, Pasley BN, Duong T, Freeman RD (2007) Transcranial magnetic stimulation elicits coupled neural and hemodynamic consequences. Science 317:1918–1921. CrossRef Medline

Anand S, Hotson J (2002) Transcranial magnetic stimulation: neurophysiological applications and safety. Brain Cogn 50:366–386. CrossRef Medline

Cohen J (1988) Statistical power analysis for the behavioral sciences. New York, NY: Routledge Academic.

Davey J, Cornelissen PL, Thompson HE, Sonkusare S, Hallam G, Smallwood J, Jefferies E (2015) Automatic and controlled semantic retrieval: TMS reveals distinct contributions of posterior middle temporal gyrus and angular gyrus. J Neurosci 35:15230–15239. CrossRef Medline

Dick AS, Mok EH, Raja Beharelle A, Goldin-Meadow S, Small SL (2014) Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. Hum Brain Mapp 35:900–917. CrossRef Medline

Drijvers L, Özyürek A (2017) Visual context enhanced: the joint contribution of iconic gestures and visible speech to degraded speech comprehension. J Speech Lang Hear Res 60:212–222. CrossRef Medline

Eickhoff SB, Laird AR, Grefkes C, Wang LE, Zilles K, Fox PT (2009) Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. Hum Brain Mapp 30:2907–2926. CrossRef Medline

Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 39:175–191. CrossRef Medline

Friederici AD (2009) Pathways to language: fiber tracts in the human brain. Trends Cogn Sci 13:175–181. CrossRef Medline

Friederici AD (2011) The brain basis of language processing: from structure to function. Physiol Rev 91:1357–1392. CrossRef Medline

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 Cogn Neurosci 18:1059–1074. CrossRef Medline

Green A, Straube B, Weis S, Jansen A, Willmes K, Konrad K, Kircher T (2009) Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. Hum Brain Mapp 30:3309–3324. CrossRef Medline

Gunter TC, Weinbrenner JE, Holle H (2015) Inconsistent use of gesture space during abstract pointing impairs language comprehension. Front Psychol 6:80. CrossRef Medline

Hadar U, Pinchas-Zamir L (2004) The semantic specificity of gesture: implications for gesture classification and function. J Lang Soc Psychol 23:204–214. CrossRef

Hagoort P (2013) MUC (Memory, Unification, Control) and beyond. Front Psychol 4:416. CrossRef Medline

Hagoort P, van Berkum J (2007) Beyond the sentence given. Philos Trans R Soc Lond B Biol Sci 362:801–811. CrossRef Medline

Hartwigsen G, Henseler I, Stockert A, Wawrzyniak M, Wendt C, Klingbeil J, Baumgaertner A, Saur D (2017a) Integration demands modulate effective connectivity in a fronto-temporal network for contextual sentence integration. Neuroimage 147:812–824. CrossRef Medline

Hartwigsen G, Bzdok D, Klein M, Wawrzyniak M, Stockert A, Wrede K, Classen J, Saur D (2017b) Rapid short-term reorganization in the language network. Elife 6:e25964. CrossRef Medline

Hayward G, Goodwin GM, Harmer CJ (2004) The role of the anterior cingulate cortex in the counting Stroop task. Exp Brain Res 154:355–358. CrossRef Medline

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402. CrossRef Medline

Holle H, Gunter TC, Rüschemeyer SA, Hennenlotter A, Iacoboni M (2008) Neural correlates of the processing of co-speech gestures. Neuroimage 39:2010–2024. CrossRef Medline

Holle H, Obleser J, Rueschemeyer SA, Gunter TC (2010) Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. Neuroimage 49:875–884. CrossRef Medline

Huang YZ, Edwards MJ, Rounis E, Bhatia KP, Rothwell JC (2005) Theta burst stimulation of the human motor cortex. Neuron 45:201–206. CrossRef Medline

Jackson RL, Hoffman P, Pobric G, Lambon Ralph MA (2016) The semantic network at work and rest: differential connectivity of anterior temporal lobe subregions. J Neurosci 36:1490–1501. CrossRef Medline

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. CrossRef Medline

Kelly SD, Creigh P, Bartolotti J (2010a) Integrating speech and iconic ges-

tures in a Stroop-like task: evidence for automatic processing. J Cogn Neurosci 22:683–694. CrossRef Medline

Kelly SD, Özyürek A, Maris E (2010b) Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. Psychol Sci 21:260–267. CrossRef Medline

Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics: (de)constructing the N400. Nat Rev Neurosci 9:920–933. CrossRef Medline

Luber B, Lisanby SH (2014) Enhancement of human cognitive performance using transcranial magnetic stimulation (TMS). Neuroimage 85:961–970. CrossRef Medline

Marstaller L, Burianová H (2014) The multisensory perception of co-speech gestures: a review and meta-analysis of neuroimaging studies. J Neurolinguistics 30:69–77. CrossRef

McNeill D, Cassell J, McCullough KE (1994) Communicative effects of speech-mismatched gestures. Res Lang Soc Interaction 27:223–237. CrossRef

Noonan KA, Jefferies E, Visser M, Lambon Ralph MA (2013) Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. J Cogn Neurosci 25:1824–1850. CrossRef Medline

Nyffeler T, Cazzoli D, Wurtz P, Lüthi M, von Wartburg R, Chaves S, Déruaz A, Hess CW, Müri RM (2008) Neglect-like visual exploration behaviour after theta burst transcranial magnetic stimulation of the right posterior parietal cortex. Eur J Neurosci 27:1809–1813. CrossRef Medline

Nyffeler T, Cazzoli D, Hess CW, Müri RM (2009) One session of repeated parietal theta burst stimulation trains induces long-lasting improvement of visual neglect. Stroke 40:2791–2796. CrossRef Medline

Oldfield RC (1971) The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9:97–113. CrossRef

Özyürek A (2014) Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. Philos Trans R Soc Lond B Biol Sci 369:20130296. CrossRef Medline

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 Cogn Neurosci 19:605–616. CrossRef Medline

Ralph MA, Jefferies E, Patterson K, Rogers TT (2017) The neural and computational bases of semantic cognition. Nat Rev Neurosci 18:42–55. CrossRef Medline

Ratcliff R (1993) Methods for dealing with reaction time outliers. Psychol Bull 114:510–532. CrossRef Medline

Ruzzoli M, Marzi CA, Miniussi C (2010) The neural mechanisms of the effects of transcranial magnetic stimulation on perception. J Neurophysiol 103:2982–2989. CrossRef Medline

Siciliano R, Hirata Y, Kelly SD (2016) Electrical stimulation over left inferior frontal gyrus disrupts hand Gesture's role in foreign vocabulary learning. Educ Neurosci 1:1–12. CrossRef

Stewart LM, Walsh V, Rothwell JC (2001) Motor and phosphene thresholds: a transcranial magnetic stimulation correlation study. Neuropsychologia 39:415–419. CrossRef Medline

Straube B, Green A, Bromberger B, Kircher T (2011) The differentiation of iconic and metaphoric gestures: common and unique integration processes. Hum Brain Mapp 32:520–533. CrossRef Medline

Straube B, Green A, Weis S, Kircher T (2012) A supramodal neural network for speech and gesture semantics: an fMRI study. PLoS One 7:e51207. CrossRef Medline

Suppa A, Huang YZ, Funke K, Ridding MC, Cheeran B, Di Lazzaro V, Ziemann U, Rothwell JC (2016) Ten years of theta burst stimulation in humans: established knowledge, unknowns and prospects. Brain Stimul 9:323–335. CrossRef Medline

Visser M, Jefferies E, Lambon Ralph MA (2010) Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. J Cogn Neurosci 22:1083–1094. CrossRef Medline

Walsh V, Ellison A, Battelli L, Cowey A (1998) Task-specific impairments and enhancements induced by magnetic stimulation of human visual area V5. Proc Biol Sci 265:537–543. CrossRef Medline

Wawrzyniak M, Hoffstaedter F, Klingbeil J, Stockert A, Wrede K, Hartwigsen G, Eickhoff SB, Classen J, Saur D (2017) Fronto-temporal interactions are functionally relevant for semantic control in language processing. PLoS One 12:e0177753. CrossRef Medline

Whitney C, Kirk M, O'Sullivan J, Lambon Ralph MA, Jefferies E (2011) The neural organization of semantic control: TMS evidence for a distributed network in left inferior frontal and posterior middle temporal gyrus. Cereb Cortex 21:1066–1075. CrossRef Medline

Willems RM, Özyürek A, Hagoort P (2007) When language meets action: the neural integration of gesture and speech. Cereb Cortex 17:2322–2333. CrossRef Medline

Willems RM, Özyürek A, Hagoort P (2009) Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. Neuroimage 47:1992–2004. CrossRef Medline

Wu YC, Coulson S (2007) How iconic gestures enhance communication: an ERP study. Brain Lang 101:234–245. CrossRef Medline

Wu YC, Coulson S (2011) Are depictive gestures like pictures? Commonalities and differences in semantic processing. Brain Lang 119:184–195. CrossRef Medline

Xu J, Gannon PJ, Emmorey K, Smith JF, Braun AR (2009) Symbolic gestures and spoken language are processed by a common neural system. Proc Natl Acad Sci U S A 106:20664–20669. CrossRef Medline

Yang J, Andric M, Mathew MM (2015) The neural basis of hand gesture comprehension: a meta-analysis of functional magnetic resonance imaging studies. Neurosci Biobehav Rev 57:88–104. CrossRef Medline

Yue Q, Zhang L, Xu G, Shu H, Li P (2013) Task-modulated activation and functional connectivity of the temporal and frontal areas during speech comprehension. Neuroscience 237:87–95. CrossRef Medline