

TMS Reveals Dynamic Interaction between Inferior Frontal Gyrus and Posterior Middle Temporal Gyrus in Gesture-Speech Semantic Integration

Wanying Zhao,¹ Yanchang Li,¹ and Yi Du^{1,2,3,4*}

¹CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China, ²Department of Psychology, University of Chinese Academy of Sciences, Beijing, 100049, China, ³CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai, 200031, China, and ⁴Chinese Institute for Brain Research, Beijing, 102206, China

Semantic processing is an amodal process with modality-specific information integrated in supramodal “convergence zones” or “semantic hub” with executive mechanisms that tailor semantic representation in a task-appropriate way. One unsolved question is how frontal control region dynamically interacts with temporal representation region in semantic integration. The present study addressed this issue by using inhibitory double-pulse transcranial magnetic stimulation over the left inferior frontal gyrus (IFG) or left posterior middle temporal gyrus (pMTG) in one of eight 40 ms time windows (TWs) (3 TWs before and 5 TWs after the identification point of speech), when human participants (12 females, 14 males) were presented with semantically congruent or incongruent gesture-speech pairs but merely identified the gender of speech. We found a TW-selective disruption of gesture-speech integration, indexed by the semantic congruency effect (i.e., a cost of reaction time because of semantic conflict), when stimulating the left pMTG in TW1, TW2, and TW7 but when stimulating the left IFG in TW3 and TW6. Based on the timing relationship, we hypothesize a two-stage gesture-speech integration circuit with a pMTG-to-IFG sequential involvement in the prelexical stage for activating gesture semantics and top-down constraining the phonological processing of speech. In the postlexical stage, an IFG-to-pMTG feedback signal might be implicated for the control of goal-directed representations and multimodal semantic unification. Our findings provide new insights into the dynamic brain network of multimodal semantic processing by causally revealing the temporal dynamics of frontal control and temporal representation regions.

Key words: gesture; inferior frontal gyrus; posterior middle temporal gyrus; semantic integration; speech; TMS

Significance Statement

Previous research has identified differential functions of left inferior frontal gyrus (IFG) and posterior middle temporal gyrus (pMTG) in semantic control and semantic representation, respectively, and a causal contribution of both regions in gesture-speech integration. However, it remains largely unclear how the two regions dynamically interact in semantic processing. By using double-pulse transcranial magnetic stimulation to disrupt regional activity at specific time, this study for the first time revealed critical time windows when the two areas were causally involved in integrating gesture and speech semantics. Findings suggest a pMTG-IFG-pMTG neurocircuit loop in gesture-speech integration, which deepens current knowledge and inspires future investigation of the temporal dynamics and cognitive processes of the amodal semantic network.

Introduction

Semantic processing, the cognitive act of accessing stored conceptual knowledge acquired from multimodal verbal and non-verbal experience, is believed to be a general process abstract away from modality-specific attributes (Caramazza et al., 1990). Convergent evidence supports that semantic cognition relies on two principal interacting neural systems (for review, see Binder et al., 2016; Ralph et al., 2017). One is the representation system, where modality-specific sensory, motor, and affective information is integrated in temporal and inferior parietal “convergence zones” (Damasio et al., 1996) and the “semantic hub” in anterior

Received June 29, 2021; revised Oct. 26, 2021; accepted Nov. 1, 2021.

Author contributions: W.Z. and Y.D. designed research; W.Z. and Y.L. performed research; W.Z. analyzed data; W.Z. wrote the first draft of the paper; Y.D. edited the paper; Y.D. wrote the paper.

The authors declare no competing financial interests.

This work was supported by National Natural Science Foundation of China Grants 31800964 and 31822024; Scientific Foundation of Institute of Psychology, Chinese Academy of Sciences Grant Y8CX382005; and Strategic Priority Research Program of Chinese Academy of Sciences Grant XDB32010300.

Correspondence should be addressed to Yi Du at duyid@psych.ac.cn.

<https://doi.org/10.1523/JNEUROSCI.1355-21.2021>

Copyright © 2021 the authors

temporal lobe that store increasing abstract concepts and knowledge (Patterson et al., 2007). The other is the control system mainly including the inferior frontal cortex (IFG), which computes and manipulates activation in the representation system to suit the current context or goals (Whitney et al., 2011; Davey et al., 2015). Yet, the time course of the interaction between those two systems in semantic processing is largely unclear, the solution of which would certainly deepen our understanding of the multisensory semantic process.

Among the multimodal extralinguistic information, gestures are of particular importance in that gestures often co-occur with speech and convey not only relevant information but also additional information not present in the accompanying speech, as pantomimes can stand on their own (Kelly and Church, 1998; Goldin-Meadow and Sandhofer, 1999). Gestures and speech are believed to interact in a bidirectional way not only in terms of the external form, such as voice spectra and gesture kinematics (Bernardis and Gentilucci, 2006), but also at the semantic level (Kelly et al., 1999; Kita and Ozyurek, 2003), with an N400 effect been triggered when gestures and speech have incongruent meanings (Kelly et al., 2004; Holle and Gunter, 2007). Neuroimaging studies have identified two areas that are consistently activated in gesture-speech integration: the left IFG, which has been implicated in controlled retrieval, selection, and unification of semantic representations; and the posterior superior temporal sulcus (pSTS)/middle temporal gyrus (MTG), which has been involved in mapping multimodal inputs onto a stored common representation (Holle et al., 2008, 2010; Dick et al., 2012). In our prior study (Zhao et al., 2018), offline theta-burst transcranial magnetic stimulation (TMS) and online repetitive TMS over the left IFG or the left posterior middle temporal gyrus (pMTG) significantly reduced the difference in reaction time (RT) between the semantically incongruent and semantically congruent gesture-speech pairs, thus providing causal evidence for the involvement of both areas in gesture-speech semantic processing. Although neurophysiology studies have uncovered an earlier activity in the pMTG than in the IFG in gesture-speech integration (Drijvers et al., 2018a; He et al., 2018), a clear picture of the dynamic interplay between the two regions in integrating spatial-motoric gestures with linear-analytic speech in semantic processing is still lacking.

The present study aimed to unpack the above question using double-pulse TMS. By targeting short time periods, double-pulse TMS can benefit from the summation effect of the two pulses in transiently inhibiting specific brain region, thus providing an ideal protocol for causally investigating the timing of the brain area in a cognitive process (Pitcher et al., 2007). Naturally, gestures precede the onset of relevant speech (Morrel-Samuels and Krauss, 1992; Holler and Levinson, 2019). It is hypothesized that gestures serve as primes in conceptualization (Kita et al., 2017) and cues to constrain the perception and lexical representation of the unfolding speech (Smith et al., 2017). Hence, the present study adopted a semantic priming paradigm (Wu and Coulson, 2007; Yap et al., 2011; So et al., 2013) by presenting the speech onset at the point when gesture started to provide a clear meaning (i.e., the discrimination point [DP] of gesture). We further targeted on the lexical retrieval and unification process of speech by segmenting the possible gesture-speech integration window (Ozyurek et al., 2007; Zhao et al., 2018) into eight 40 ms time windows (TWs) surrounding the identification point (IP) of speech (i.e., the first time point when speech becomes semantically identified). We then applied double-pulse TMS in each TW to shortly inhibit the left IFG, the left pMTG, or the vertex (served as a control site) in a time-locked manner (Fig. 1A). By

doing so, we could provide a direct investigation into the time courses and functional roles the left IFG and the left pMTG may play in gesture-speech semantic integration if a dynamic interplay between the two regions indeed existed.

Materials and Methods

Participants. A total of 148 native Chinese speakers signed written informed consent forms approved by the Institute of Psychology, Chinese Academy of Sciences, and took part in the experiment. Thirty-two subjects (17 females, aged 19–28 years, SD = 8.19 years) participated in Pretest 1 for semantic congruency rating. Thirty participants (17 females, aged 19–28 years, SD = 9.92 years) took part in Pretest 2 to validate the stimuli. Another 2 sets of 30 subjects participated in Pretest 3 (16 females, aged 19–35 years, SD = 11.03 years) and Pretest 4 (15 females, aged 19–30 years, SD = 10.26 years) for the gating experiments of gesture and speech stimuli. Twenty-six participants (12 females, aged 18–28 years, mean = 22, SD = 7.83) who did not participate in any pretest took part in the TMS study. All participants were right-handed, had normal hearing, had normal or corrected-to-normal vision, and were paid ¥100 per hour for their participation.

Stimuli. The stimuli were revised based on our previous study (Zhao et al., 2018). Forty-four original, English common action videos were selected and translated into Mandarin Chinese, resulting in a total of 28 qualified actions. Two native Chinese speakers (1 male, 1 female) produced each action while uttering the corresponding Chinese word. The Chinese audios were then combined with the relevant videos produced by English speakers to generate the two experimental manipulations, the gender congruency factor (e.g., a man doing a gesture combined with a male voice or a woman doing a gesture combined with a male voice) and the semantic congruency factor (e.g., a man or a woman doing a “cut” gesture while saying Mandarin speech word “剪jian3 (cut)” or a man or a woman doing a “spray” gesture while saying “喷pen1 (cut)”). To counterbalance across all the stimulus sets, the reverse combination was also used (e.g., a man or a woman doing a “cut” gesture while saying “喷pen1 (spray)”) (for details, see Zhao et al., 2018). A total of 56 pairs of gestures and speech were used in the following four pretests.

Pretest 1: semantic congruency rating. To verify that the semantically congruent or incongruent combinations of gestures and speech were indeed perceived as such, 32 participants rated the relationship between the 56 pairs of gestures and speech on a 5-point rating scale (1 means “no relation” and 5 means a “very strong relation”). Based on the rating results, eight pairs of stimuli were moved from the main stimulus set to the practice set as examples of ambiguous relationship. The 48 remaining stimuli were used for further pretests. The mean rating for the remaining set of congruent pairs was 4.48 (SD = 0.40), and the mean rating for the incongruent pairs was 1.44 (SD = 0.42).

Pretest 2: stimulus set validation. Another 30 participants validated the stimulus set by replicating the semantic congruency effect in previous studies (Kelly et al., 2010). Participants were informed that the gender they saw on the screen and the gender of the voice they heard might be different or the same. Participants were asked to merely pay attention to the gender of the voice they heard and press one of two buttons on the keyboard (key “F” for male and key “J” for female in half of participants; reversed in the other half) as quickly and accurately as possible. The video started at the stroke of gestures. Speech onset occurred 200 ms after the onset of the video. RT was recorded relative to the onset of speech, and participants pressed the button at the moment of video stop.

Participants made very few errors in the task (overall accuracy = 96.49%); therefore, accuracy data were not analyzed, and incorrect trials were excluded. The remaining correct responses were trimmed within 2 SDs of each subject’s mean RT. Overall, this resulted in 4.2% of trials being excluded as outliers. To maximize the semantic congruency effect, eight pairs of stimuli were further deleted. The rest 40 pairs of stimuli constituted 2 (Semantic congruency) × 2 (Gender congruency) conditions, with 160 trials in total. Similar to previous findings, a 2 (Semantic congruency) × 2 (Gender congruency) repeated-measures ANOVA revealed a significant main effect of Gender congruency ($F_{(1,29)} = 45.46$, $p < 0.001$, $\eta_p^2 = 0.611$), with gender-incongruent

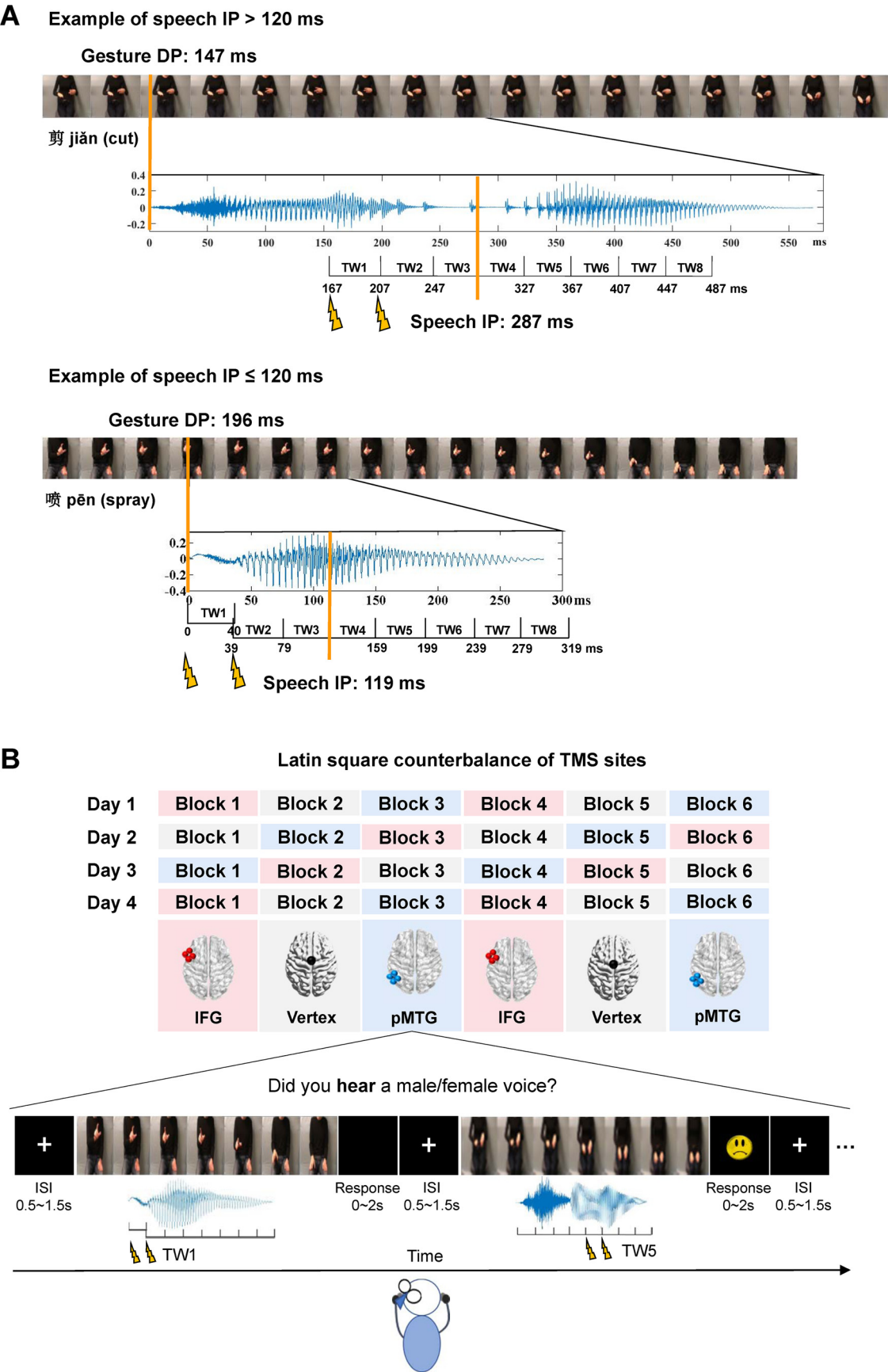


Figure 1. Stimuli and experimental design. **A**, Illustration of the 8 TWs (duration = 40 ms) relative to the IP of speech in which TMS was applied. Speech onset coincided with the DP of the gesture. Top, Example of a speech IP >120 ms. Bottom, Example of a speech IP ≤120 ms. **B**, A total of 1280 gesture–speech pairs were split into 24 blocks that were completed on 4 different days with 5–7 d apart between each other. There were 8 blocks for each stimulation site, and the order of blocks was counterbalanced using a Latin square across participants. In each block,

Table 1. The mean DP of each action gesture and mean IP of each speech word

Item	DP_female (ms)	DP_male (ms)	DP_mean (ms)	IP_female (ms)	IP_male (ms)	IP_mean (ms)
an4 (press)	164.14	169.33	167	158.62	125.52	142
bai1 (break)	121.38	95.71	109	213.33	122.86	168
ca1 (wipe)	111.72	130.00	121	200.00	146.67	173
ding1 (hammer)	155.71	148.57	152	182.22	84.14	133
feng2 (sew)	313.33	247.14	280	248.28	277.24	263
he2 (close)	242.76	350.77	297	84.29	84.29	84
ji3 (squeeze)	118.57	117.14	118	151.72	162.76	157
jian3 (cut)	169.66	124.29	147	277.33	296.00	287
jiao3 (stir)	201.38	283.33	242	297.14	180.69	239
la1 (pull)	152.86	100.69	127	113.10	91.85	102
ning2 (twist)	118.62	85.93	102	271.72	175.71	224
pen1 (spray)	238.67	175.17	207	114.48	122.86	119
reng1 (throw)	161.38	117.04	139	161.38	113.10	137
shan1 (slap)	94.29	94.29	94	291.85	234.07	263
shua1 (scrub)	413.33	357.24	385	278.67	191.43	235
si1 (tear)	100.74	238.57	170	147.59	128.28	138
ti2 (lift)	105.19	122.96	114	121.38	100.74	111
tui1 (push)	241.38	157.14	199	257.93	161.33	210
yao2 (shake)	204.29	157.14	199	196.00	158.62	177
yun4 (iron)	295.17	318.67	307	206.90	124.00	165

trials (mean = 556.64 ms, SE = 12.11) eliciting slower RTs than gender-congruent trials (mean = 531.78 ms, SE = 11.67). Importantly, a significant main effect of Semantic congruency ($F_{(1,29)} = 51.12$, $p < 0.001$, $\eta_p^2 = 0.638$) was replicated: participants were slower to judge the gender of the speaker when speech and gesture were semantically incongruent (mean = 554.51 ms, SE = 11.65) relative to when they were semantically congruent (mean = 533.90 ms, SE = 12.02). However, no significant interaction of Semantic congruency and Gender congruency was found ($F_{(1,29)} = 0.542$, $p = 0.468$, $\eta_p^2 = 0.018$).

Pretest 3: gating study of gesture stimuli. We used the gating paradigm (Obermeier et al., 2011; Obermeier and Gunter, 2015) to define the minimal length of each gesture required for semantic identification, namely, the DP of gesture (Table 1). To do so, the remaining 20 gestures (length = 1771.00 ms, SD = 307.98 ms) performed by either a male or a female were presented to 30 participants. Each gesture was presented without speech and in segments of increasing duration at a step of 40 ms (the first segment started at 40 ms). Participants were told that they would be presented with a number of videos of someone performing an action without holding the object. For each action, there were several videos of various durations. Participants were asked to infer what was described in the action with a single action word. There was no time limit in which participants had to respond. Participants could move to the next action by either giving no correct answer at the end of the action or giving the same answer for that action 3–6 times continuously (time varied to prevent a learning effect). The DP of a gesture was defined as the first time point the participant gave the final answer.

To eliminate the influence of outliers, RTs outside 2 SDs of the mean for each gesture were excluded (5.5% of trials). On average, the DP of gestures was 183.78 ms (SD = 84.82 ms). Paired-sample t tests showed that there was no significant difference in whether the action was performed by a male or a female ($t_{(20)} = 0.21$, $p = 0.84$).

Pretest 4: gating study of speech stimuli. To locate the IP of speech, the 20 action verbs pronounced by either a male or a female speaker (length = 447.08 ms, SD = 93.48 ms) were presented to another set of 30 participants. Each sound was presented in segments of increasing duration that started at a length of 80 ms, with an increase of 40 ms.

Participants were asked to listen carefully, try to infer what was presented in the audio, and write down the word they heard. All other procedures were the same as those used in Pretest 3. Similar to the DP of gestures, the IP of speech was defined as the first time point participants gave the final answer of the various speech segments presented.

After removing the outliers (> 2 SDs of the mean, 4.6% of trials) for each speech, on average, the IP of speech was 176.40 ms (SD = 66.21 ms, Table 1). Paired-sample t tests showed that there was no significant difference in whether the speech was pronounced by a male or a female ($t_{(20)} = 0.52$, $p = 0.61$).

Experimental procedure. We used double-pulse TMS to investigate the temporal specificity of the involvement of the left IFG and the left pMTG in gesture-speech integration. Previous studies have shown that double pulses with a pulse interval of 40 ms were enough to elicit a transient time-locked “virtual lesion” to the normal neural firing pattern and generate an inhibitory effect on cortical functions (O’Shea et al., 2004; Pitcher et al., 2007, 2008). Therefore, the present study applied double-pulse TMS at the boundaries of each of the 40 ms TWs over either the left IFG or the left pMTG to examine the precise timing of two regions in gesture-speech integration. Notably, a mixed TMS effect between stimulation sites was not considered because of the lack of aftereffect of double-pulse TMS and the balanced order of stimulation sites across participants.

Eight 40 ms TWs were segmented relative to the speech IP. There were 3 TWs before the speech IP (TW1: –120 to 80 ms; TW2: –80 to 40 ms; and TW3: –40 to 0 ms) and 5 TWs after the speech IP (TW4: 0–40 ms; TW5: 40–80 ms; TW6: 80–120 ms; TW7: 120–160 ms; and TW8: 160–200 ms). To ensure that all TWs were located after the onset of speech, for those action verbs with IP < 120 ms, the first TW was defined as from the onset of speech to 40 ms after onset (Fig. 1A).

To eliminate effect caused by stimuli, each of the 160 gesture-speech pairs underwent TMS in each of the 8 TWs, leading to 1280 trials in total. The 1280 trials were further split into 24 blocks (54 trials in the first 20 blocks, and 50 trials in the last 4 blocks) with 8 blocks for each stimulation site, and the order of blocks was counterbalanced using a Latin square design across participants. Participants completed the 24 blocks on 4 different days that were 5–7 d apart to avoid fatigue and the learning effect. In each block, one area was stimulated by double-pulse TMS in all 8 TWs in a random order (Fig. 1B). In each trial, a fixation cross was first presented on the center of the screen for 0.5–1.5 s, followed by a video of gesture and speech, and participants were asked to look at the screen but merely respond to the gender of the voice within 2 s. Feedback was given only if the response was incorrect. RT was recorded from the onset of speech. A 4 (Day) $\times$ 2 (Semantic congruency) $\times$ 2 (Gender congruency) repeated-measures ANOVA revealed a significant main effect of Day

←

double-pulse TMS was conducted on one stimulation site [the left IFG (red spot), the left pMTG (blue spot), or the vertex (black spot)] in all 8 TWs randomly. In each trial, a fixation cross was first presented on the center of the screen for 0.5–1.5 s, followed by a video of gesture and speech, and participants were asked to respond to the gender of the voice within 2 s. A feedback was given only if the response was incorrect.

($F_{(1.648,41.192)} = 19.590$, $p < 0.001$, $\eta_p^2 = 0.439$), with gradually decreased RTs as participants became more familiar with the task. However, there was no modulation of Day on either Semantic congruency ($F_{(2.031,50.775)} = 2.110$, $p = 0.131$, $\eta_p^2 = 0.078$) or Gender congruency ($F_{(2.172,54.300)} = 0.532$, $p = 0.605$, $\eta_p^2 = 0.021$).

The stimuli were presented using Presentation software (version 17.2, www.neurobs.com). All other procedures were the same as those described for Pretest 2. Before the formal experiment, participants performed 16 training trials to become familiar with the experimental procedure.

TMS protocol. The stimulation sites of the left IFG (−62, 16, 22) and the left pMTG (−50, −56, 10) corresponding to MNI coordinates were identified in a quantitative meta-analysis of fMRI studies on iconic gesture-speech integration (for details, see Zhao et al., 2018). The vertex was used as a control site.

To enable image-guided TMS navigation, high-resolution ($1 \times 1 \times 0.6$ mm) T1-weighted anatomic MRI scans of each participant were acquired at the Beijing MRI Center for Brain Research using a Siemens 3T Trio/Tim Scanner. Frameless stereotaxic procedures (BrainSight 2; Rogue Research) were used for online checking of stimulation during navigation. To ensure precise stimulation of each target region in each participant, individual anatomic images were manually registered by identifying the anterior and posterior commissures. Subject-specific target regions were defined by trajectory markers using the MNI coordinate system. The angles of the markers were checked and adjusted to be orthogonal to the skull during neuronavigation.

A Magstim Rapid² stimulator (Magstim) was used to deliver the double-pulse TMS via a 70 mm figure-eight coil. The double-pulse TMS at an intensity of 50% of maximum stimulator output was delivered “online” in each TW. For example, in a trial where stimulation took place in TW1, a participant would receive a pulse at time −120 ms (relative to the speech IP) and a second pulse 40 ms later (−80 ms of the speech IP). All other apparatus details were the same as those described for Pretest 2.

Data analyses. All incorrect responses (1471 out of the total number of 33,280, 4.42% of trials) were excluded. To eliminate the influence of outliers, a 2 SD trimmed mean of RT for every participant in each session was conducted. First, we conducted a 3 (Site) $\times$ 2 (Semantic congruency) $\times$ 2 (Gender congruency) repeated-measures ANOVA to examine the general effects of Site, Semantic congruency, and Gender congruency on RTs. Next, we implemented an 8 (TW) $\times$ 2 (Semantic congruency) repeated-measures ANOVA on the vertex condition alone to ensure that vertex stimulation in different TWs did not significantly change RT or the semantic congruency effect. A similar 8 (TW) $\times$ 2 (Gender congruency) ANOVA on the vertex condition was conducted, too. By doing so, we could average RTs across TWs in the vertex condition to serve as the RT baselines of the four stimulus conditions to further test the TMS effect.

Then, we focused our analysis on the TMS effect (active-TMS – vertex-TMS) and its interaction with TW over the semantic congruency factor to determine in which TW the semantically congruent condition and the semantically incongruent condition were differently affected when activity in the IFG or pMTG was disrupted relative to vertex stimulation. Accordingly, we implemented a 2 (Site: pMTG-vertex, IFG-vertex) $\times$ 8 (TW) repeated-measures ANOVA directly on the semantic congruency effect ($RT_{\text{semantically incongruent}} - RT_{\text{semantically congruent}}$), followed by one-sample t tests with false discovery rate (FDR) correction to identify TWs in which the semantic congruency effect was significantly disrupted by TMS on each site.

We also used the TMS effect over the factor of gender congruency as a control, with the assumption that double-pulse TMS would selectively impact on semantic congruency but not on gender congruency, as reflected by an insignificant two-way interaction measured by a 2 (Site) $\times$ 8 (TW) repeated-measures ANOVA. In all ANOVAs,

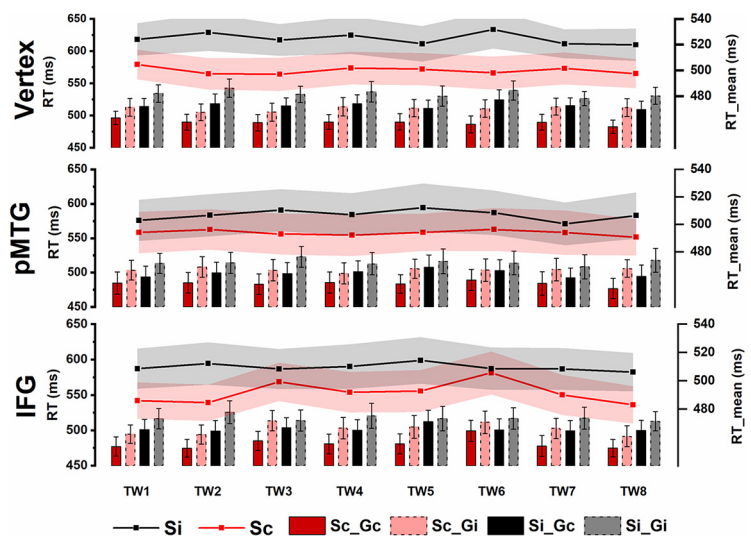


Figure 2. RTs of four experimental conditions (bar graphs) and the semantic congruent and incongruent conditions (line graphs) for the three stimulation sites. Sc, Semantically congruent; Si, semantically incongruent; Gc, gender congruent; Gi, gender incongruent. RT_{mean} represents the mean RT for the Sc and Si conditions. Error bars in the bar graphs and shadows in the line graphs indicate SEM.

Greenhouse–Geisser adjustment was applied to correct for violations of the sphericity assumption where necessary; multiple comparisons were corrected by Bonferroni correction.

Results

RTs at each experimental condition are illustrated in Figure 2. RT was generally longer in the vertex condition (mean = 512.65 ms, SE = 12.43 ms) than in the pMTG (mean = 500.55 ms, SE = 14.93 ms) and IFG (mean = 500.90 ms, SE = 13.96 ms) stimulation conditions. However, the difference was not significant, as a 3 (Site) $\times$ 2 (Semantic congruency) $\times$ 2 (Gender congruency) repeated-measures ANOVA revealed an insignificant main effect of Site ($F_{(1.831,45.778)} = 0.944$, $p = 0.396$, $\eta_p^2 = 0.036$). Consistent with previous studies (Kelly et al., 2010; Zhao et al., 2018), there was a significant main effect of Semantic congruency ($F_{(1,25)} = 255.40$, $p < 0.001$, $\eta_p^2 = 0.911$), with longer RTs in semantically incongruent trials (mean = 513.97 ms, SE = 12.80 ms) than in congruent trials (mean = 495.44 ms, SE = 12.31 ms). There was also a significant main effect of Gender congruency ($F_{(1,25)} = 71.00$, $p < 0.001$, $\eta_p^2 = 0.740$), with longer RTs when speech and gestures were produced by conflicting genders (mean = 514.14 ms, SE = 12.98 ms) than when they were produced by the same gender (mean = 495.27 ms, SE = 12.20 ms). Furthermore, there was no significant interaction between Site and Semantic congruency ($F_{(1.940,48.509)} = 2.93$, $p = 0.065$, $\eta_p^2 = 0.105$), nor between Site and Gender congruency ($F_{(1.951,48.785)} = 0.37$, $p = 0.69$, $\eta_p^2 = 0.015$).

An 8 (TW) $\times$ 2 (Semantic congruency) repeated-measures ANOVA on vertex data revealed neither a main effect of TW ($F_{(5.126,128.152)} = 0.992$, $p = 0.439$, $\eta_p^2 = 0.038$) nor an interaction of TW by Semantic congruency ($F_{(5.556,138.907)} = 2.03$, $p = 0.07$, $\eta_p^2 = 0.075$). Another ANOVA of 8 (TW) $\times$ 2 (Gender congruency) showed similar effects, neither a main effect of TW ($F_{(5.096,127.410)} = 1.049$, $p = 0.399$, $\eta_p^2 = 0.040$) nor a TW $\times$ Gender congruency interaction ($F_{(4.796,119.897)} = 0.359$, $p = 0.869$, $\eta_p^2 = 0.014$) was found. These results suggest that double-pulse TMS on the vertex in different TWs did not significantly change RT, or the semantic or gender congruency effect. To simplify analyses and get reliable measures, we averaged RTs across TWs in the vertex condition to generate

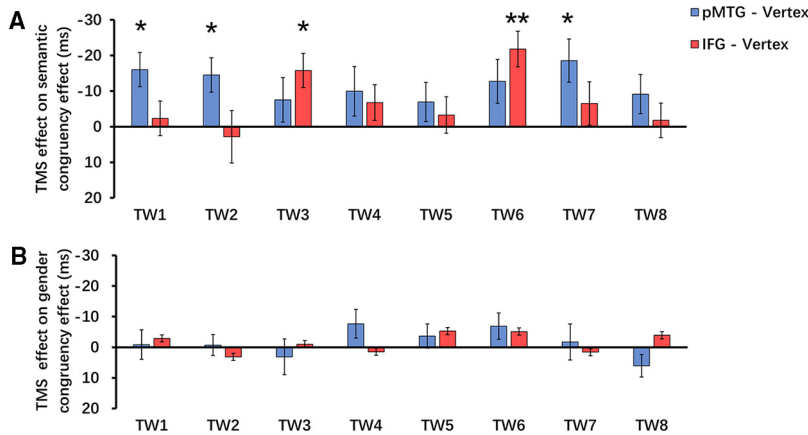


Figure 3. TMS effects on the semantic congruency effect (**A**) and the gender congruency effect (**B**). TMS effect was defined as active-TMS minus vertex-TMS. The semantic congruency effect was calculated as the RT difference between semantically incongruent and semantically congruent pairs, whereas the gender congruency effect was calculated as the RT difference between gender incongruent and gender congruent pairs. * $p < 0.05$; ** $p < 0.01$; one-sample t tests after FDR correction. Error bars represent SEM.

the baseline RTs of the four stimulus conditions in examining the TMS effect.

To directly show the Site- and TW-specific TMS effect on the semantic congruency effect (i.e., an RT cost because of semantic conflict), a 2 (Site: pMTG-vertex, IFG-vertex) $\times$ 8 (TW) repeated-measures ANOVA on the semantic congruency effect ($RT_{\text{semantically incongruent}} - RT_{\text{semantically congruent}}$) revealed a significant Site $\times$ TW interaction ($F_{(5,247,131,175)} = 2.252, p = 0.034, \eta_p^2 = 0.083$) (Fig. 3A). Following one-sample t tests showed a significant TMS disruption of the semantic congruency effect (negative values in Fig. 3A indicate decreased semantic congruency effect compared with the baseline) when stimulating the pMTG in TW1 ($t_{(25)} = 3.337$, FDR-corrected $p = 0.015$, Cohen's $d = 0.645$), TW2 ($t_{(25)} = 3.019$, FDR-corrected $p = 0.015$, Cohen's $d = 0.592$), and TW7 ($t_{(25)} = 3.063$, FDR-corrected $p = 0.015$, Cohen's $d = 0.601$). Similar TMS impairment of the semantic congruency effect was found when stimulating the IFG in TW3 ($t_{(25)} = 3.299$, FDR-corrected $p = 0.012$, Cohen's $d = 0.647$) and TW6 ($t_{(25)} = 4.348$, FDR-corrected $p = 0.002$, Cohen's $d = 0.853$). To further depict how the semantically congruent condition and the semantically incongruent condition were differentially affected by TMS stimulation, the TMS effects on each condition were shown in Figure 4. The significant TMS disruption of the semantic congruency effect when stimulating the pMTG in TW1, TW6, and TW7 was caused mainly by a TMS-induced decrease of RT in the semantically incongruent condition without influencing the semantically congruent condition. In contrast, the significant TMS impairment of the semantic congruency effect when stimulating the IFG in TW3 and TW6 was because of a TMS-induced increase of RT in the semantically congruent condition without impacting the semantically incongruent trials.

As a control analysis, another repeated-measures ANOVA of 2 (Site) $\times$ 8 (TW) on the gender congruency effect showed an insignificant two-way interaction ($F_{(4,999,124,970)} = 0.750, p = 0.587, \eta_p^2 = 0.029$), indicating no TW-specific TMS modulation on gender congruency (Fig. 3B).

Discussion

By splitting the integration process of gestures and speech into 8 TWs, and applying double-pulse TMS in each of the TWs on

either the left IFG or the left pMTG, brain areas considered as the neural underpinnings supporting semantic control and cross-modal representations, respectively, during gesture-speech integration (Willems et al., 2009; Zhao et al., 2018), we created a novel paradigm to investigate how the left IFG and the left pMTG temporally interact in gesture-speech semantic processing. Crucially, our results, for the first time, revealed a causal involvement of both areas with differential time courses in automatic semantic processing. As summarized in Figure 5A, when speech had not reached its semantic IP, the semantic congruency effect (i.e., the RT cost induced by semantic conflict between gestures and speech) was significantly disrupted by double-pulse TMS over the left pMTG in TW1 and TW2 and over the left IFG in TW3 relative to vertex stimulation. After the speech reaching its semantic IP, significant TMS impairment of the semantic congruency effect was found when conducting TMS on the left IFG in TW6 and over the left pMTG in TW7. Our findings provide causal evidence for a sequential engagement in gesture-speech semantic integration from the left pMTG to the left IFG at the prelexical stage and from the left IFG to the left pMTG at the postlexical stage, a two-stage gesture-speech integration circuit was thus proposed in Figure 5B. As an extension of our previous findings (Zhao et al., 2018) that both the left pMTG and the left IFG played a causal contribution to semantic integration of gestures and speech, the present study fills the gap of knowledge in understanding how the frontal control region and temporal storage node dynamically interplay in integrating multimodal semantic information.

Bidirectional neural pathways connecting the left IFG and pMTG in semantic processing have been proposed extensively (Hickok and Poeppel, 2007; Hagoort, 2013; Friederici et al., 2017). For instance, the Memory-Unification-Control model (Hagoort, 2013) proposed a recurrent neurocircuit connecting the left IFG and pMTG for semantic unification. In this circuit, semantic features and lexical information are activated in posterior temporal regions and relayed to the IFG; then IFG neurons send feedback signals to the temporal cortex to manipulate the activation level of semantic representations, so as to maintain the context and unify the lexical information within the context. Using double-pulse TMS with high temporal resolution, the present study offers the first causal evidence for the dynamic interplay between the left IFG and the left pMTG, which echoes such a recurrent neurocircuit in the sense of multimodal semantic processing. Specifically, we observed a pMTG-to-IFG timing sequence in the stage of mapping acoustic speech input onto corresponding lexical-conceptual semantic representation, as indexed by the involvement of the left pMTG in TW1 and TW2 (120–40 ms before speech IP) and the left IFG in TW3 (40–0 ms before speech IP). There was also an IFG-to-pMTG timing relationship after the semantic retrieval of speech, as shown by the involvement of the left IFG in TW6 (80–120 ms after speech IP) and the left pMTG in TW7 (120–160 ms after speech IP).

Naturalistically, the multisensory information is not strictly aligned in time and the process of integration is supposed to take place at various stages of bottom-up and top-down cortical

interplays (for review, see Talsma, 2015; Xu et al., 2020). By presenting the onset of speech at the DP of gesture, the present study created a semantic priming paradigm of gestures onto speech. We therefore comprehend the results as a crosstalk of the top-down modulation of gestures with the bottom-up processing of speech. In the first prelexical stage, semantic information extracted from the gesture would constrain the phonological encoding of speech, while in the second postlexical stage, the most feasible lexical candidate was selected, retrieved, and unified with the gesture semantics to form context-appropriate semantic representations.

Furthermore, there seems a division of labor of the left pMTG and the left IFG in each stage of gesture-speech integration. Although TMS on both areas disrupted the semantic congruency effect, stimulation of the pMTG led to shorter RT in the semantically incongruent condition, whereas stimulation of the IFG led to longer RT in the semantically congruent condition (see Fig. 4). It is believed that the pMTG is involved in mediating the long-term storage of, and access to, supramodal semantic representations, while the left IFG has been reliably associated with controlled retrieval and selection of lexical representations to fit the current context or goal, and this process is independent of modality (for review, see Lau et al., 2008; Binder and Desai, 2011; Ralph et al., 2017). Consistent with those claims, we interpret the functional roles of the two regions in the proposed two-stage gesture-speech integration circuit (Fig. 5B) as follows. In the first stage, disrupting activity in the left pMTG may preclude both the top-down modulation of semantic information extracted from the gesture on the lower-level phonological processing of speech in the superior temporal gyrus and sulcus (STG/STS) (Bizley et al., 2016), and the bottom-up mapping of the phonological speech to lexical representation in the pMTG. This would likely lead to a failure in monitoring the semantic conflict between gestures and speech, as reflected by decreased RT only in semantically incongruent trials. In the second stage, although speech started at the point when gestures became semantically clear, we cannot dismiss that perturbing the pMTG activity would interfere with the reanalysis of semantic information for the observed iconic gesture to make it compatible with the accompanying speech context (Fritz et al., 2021). Since semantically incongruent pairs triggered an increased need for strategic recovery, inhibition of the pMTG in the second stage would dampen such a process, thus reducing the RT cost in the semantically incongruent condition. In contrast, disturbing activity in the left IFG likely impeded the controlled selection of the most appropriate

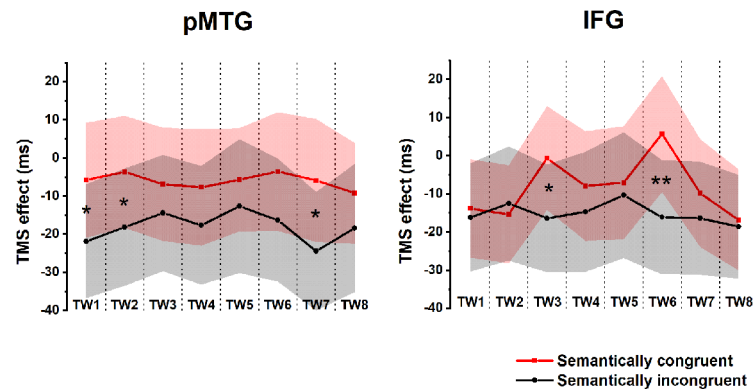


Figure 4. TMS effects on semantically congruent and incongruent conditions. TMS effects, defined as active-TMS minus vertex-TMS, on RTs of the semantically congruent (red) and semantically incongruent (black) conditions are shown for stimulating the left pMTG and left IFG in 8 TWs. * $p < 0.05$; ** $p < 0.01$; one-sample t tests after FDR correction. Shadows represent SEM.

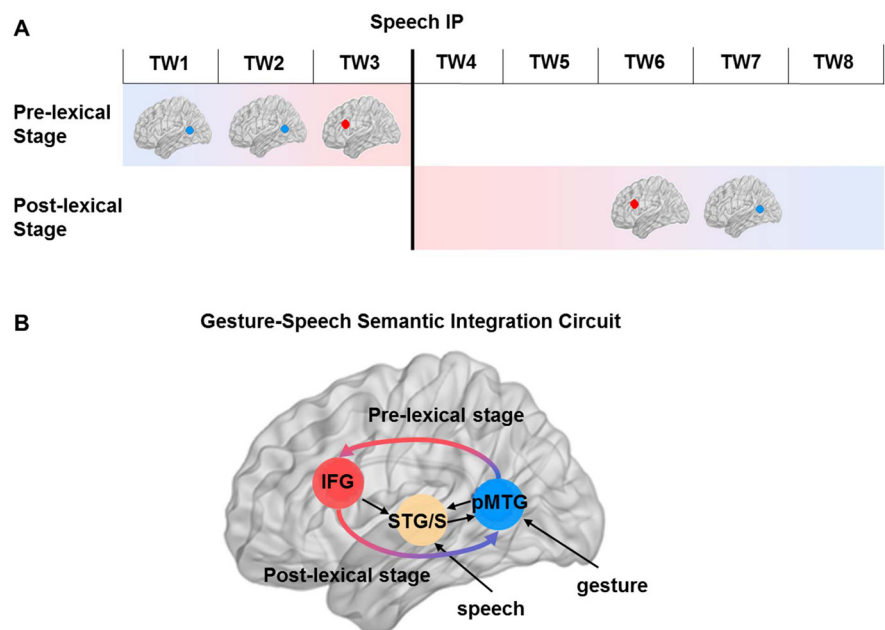


Figure 5. Summary of results and the proposed two-stage integration circuit in semantic processing of gestures and speech. **A**, Summary of critical time periods that the left IFG (red spot) and left pMTG (blue spot) get involved in gesture-speech semantic integration. **B**, A gesture-speech semantic integration circuit was posited in which there were a pMTG-to-IFG sequential involvement at the prelexical stage (before speech reached its IP) for accessing and activating gesture semantics and top-down modulation of phonological processing in the STG/STS, and an IFG-to-pMTG timing relationship at the postlexical stage (after speech IP) for controlled retrieval and multimodal semantic unification.

lexical semantics with the current context, which may impact the top-down constraining of phonological processing in the STG/STS in the first stage, and the manipulation and unification of context-appropriate supramodal semantic representations in the pMTG in the second stage. Therefore, the TMS effect was reflected as increased RT in the semantically congruent condition but no substantial effect in the semantically incongruent condition following IFG stimulation at both stages. The stimulation effect on both sites was specific to semantic processing but not general cognitive processing because no effect on the task-relevant gender congruency effect was observed.

The two-stage gesture-speech integration loop proposed here is consistent with recent findings concerning gesture and speech processing. In a simultaneous EEG and fMRI study (He et al., 2018), an α -pSTS/MTG correlation was found in an earlier TW

of gesture-speech integration and an α -IFG correlation was found in a later TW of integration. Similarly, an MEG study revealed an early α power suppression in right STS and a late α suppression in left IFG when gestures semantically disambiguated degraded speech comprehension (Drijvers et al., 2018b). Go beyond the prior knowledge, the current study made use of double-pulse TMS and offered straightforward proof for distinct engagement and temporal dynamics of the left IFG and the left pMTG in dichotomous stages of gesture-speech semantic processing.

Nonetheless, the conclusion of such a neurocircuit needs to be drawn with caution. The fact that TMS not only affects activity of the perturbed brain area but also influences activity of areas that are functionally connected with the perturbed brain area (Jackson et al., 2016; Hartwigsen et al., 2017) makes the cause-effect relationship of TMS stimulation with behavioral performance much more complex than previously thought (Bergmann and Hartwigsen, 2021). On one hand, the present results cannot tell whether the sequential involvement of the left pMTG and the left IFG was because of parallel modulations of the frontal cortex (Jacklin et al., 2016) and the temporal cortex (Noesselt et al., 2007) over the phonological processing of speech, or the information flow from the pMTG to the IFG before acting on the auditory region. On the other hand, whether there are other brain areas, such as the primary motor cortex (Marco et al., 2015) and the anterior temporal lobe (Patterson et al., 2007), that sequentially modulate activity in the pMTG and the IFG and involve in gesture-speech integration should also be clarified in future. Follow-up studies are encouraged to combine neuroimaging measures with TMS to truly unravel the functional roles and dynamic interaction of the IFG and the pMTG as well as the rapid reorganization of a wider semantic network in gesture-speech integration.

In conclusion, by applying double-pulse TMS covering the whole integration process, the present study is the first to provide causal evidence for the temporal dynamics of the left IFG and the left pMTG in gesture-speech integration. Our findings suggest a two-stage gesture-speech semantic integration circuit. In the early prelexical stage, semantic information extracted from gestures may exert top-down modulation over the phonological processing of speech, with left pMTG acting ahead of the left IFG. In the late post-lexical stage, speech may unify with gestures to form context-appropriate semantic representations through a feedback signal from the left IFG to the left pMTG. This study promises for a further understanding of the dynamic interaction between the frontal control and temporal representation regions in multimodal semantic processing.

References

- Bergmann TO, Hartwigsen G (2021) Inferring causality from noninvasive brain stimulation in cognitive neuroscience. *J Cogn Neurosci* 33:195–225.
- Bernardis P, Gentilucci M (2006) Speech and gesture share the same communication system. *Neuropsychologia* 44:178–190.
- Binder JR, Desai RH (2011) The neurobiology of semantic memory. *Trends Cogn Sci* 15:527–536.
- Binder JR, Conant LL, Humphries CJ, Fernandino L, Simons SB, Aguilar M, Desai RH (2016) Toward a brain-based componential semantic representation. *Cogn Neuropsychol* 33:130–174.
- Bizley JK, Maddox RK, Lee AK (2016) Defining auditory-visual objects: behavioral tests and physiological mechanisms. *Trends Neurosci* 39:74–85.
- Caramazza A, Hillis AE, Rapp BC, Romani C (1990) The multiple semantics hypothesis: multiple confusions? *Cogn Neuropsychol* 7:161–189.
- Damasio H, Grabowski TJ, Tranel D, Hichwa RD, Damasio AR (1996) A neural basis for lexical retrieval. *Nature* 380:499–505.
- Davey J, Rueschemeyer SA, Costigan A, Murphy N, Krieger-Redwood K, Hallam G, Jefferies E (2015) Shared neural processes support semantic control and action understanding. *Brain Lang* 142:24–35.
- Dick AS, Goldin-Meadow S, Solodkin A, Small SL (2012) Gesture in the developing brain. *Dev Sci* 15:165–180.
- Drijvers L, Ozyurek A, Jensen O (2018a) Alpha and beta oscillations index semantic congruency between speech and gestures in clear and degraded speech. *J Cogn Neurosci* 30:1086–1097.
- Drijvers L, Ozyurek A, Jensen O (2018b) Hearing and seeing meaning in noise: alpha, beta, and gamma oscillations predict gestural enhancement of degraded speech comprehension. *Hum Brain Mapp* 39:2075–2087.
- Friederici AD, Chomsky N, Berwick RC, Moro A, Bolhuis JJ (2017) Language, mind and brain. *Nat Hum Behav* 1:713–722.
- Fritz I, Kita S, Littlemore J, Krott A (2021) Multimodal language processing: how preceding discourse constrains gesture interpretation and affects gesture integration when gestures do not synchronise with semantic affiliates. *J Mem Lang* 117:104191.
- Goldin-Meadow S, Sandhofer CM (1999) Gestures convey substantive information about a child's thoughts to ordinary listeners. *Dev Sci* 2:67–74.
- Hagoort P (2013) MUC (Memory, Unification, Control) and beyond. *Front Psychol* 4:416.
- Hartwigsen G, Bzdok D, Klein M, Wawrzyniak M, Stockert A, Wrede K, Classen J, Saur D (2017) Rapid short-term reorganization in the language network. *Elife* 6:e25964.
- He YF, Steines M, Sommer J, Gebhardt H, Nagels A, Sammer G, Kircher T, Straube B (2018) Spatial-temporal dynamics of gesture-speech integration: a simultaneous EEG-fMRI study. *Brain Struct Funct* 223:3073–3089.
- Hickok G, Poeppel D (2007) The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Holle H, Gunter TC (2007) The role of iconic gestures in speech disambiguation: ERP evidence. *J Cogn Neurosci* 19:1175–1192.
- Holle H, Gunter TC, Rueschemeyer SA, Hennenlotter A, Iacoboni M (2008) Neural correlates of the processing of co-speech gestures. *Neuroimage* 39:2010–2024.
- 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.
- Holler J, Levinson SC (2019) Multimodal language processing in human communication. *Trends Cogn Sci* 23:639–652.
- Jacklin DL, Cloke JM, Potvin A, Garrett I, Winters BD (2016) The dynamic multisensory engram: neural circuitry underlying crossmodal object recognition in rats changes with the nature of object experience. *J Neurosci* 36:1273–1289.
- Jackson RL, Hoffman P, Pobric G, Ralph MA (2016) The semantic network at work and rest: differential connectivity of anterior temporal lobe subregions. *J Neurosci* 36:1490–1501.
- Kelly SD, Church RB (1998) A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Dev* 69:85–93.
- Kelly SD, Barr DJ, Church RB, Lynch K (1999) Offering a hand to pragmatic understanding: the role of speech and gesture in comprehension and memory. *J Mem Lang* 40:577–592.
- Kelly SD, Kravitz C, Hopkins M (2004) Neural correlates of bimodal speech and gesture comprehension. *Brain Lang* 89:253–260.
- Kelly SD, Creigh P, Bartolotti J (2010) Integrating speech and iconic gestures in a Stroop-like task: evidence for automatic processing. *J Cogn Neurosci* 22:683–694.
- Kita S, Ozyurek A (2003) What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *J Mem Lang* 48:16–32.
- Kita S, Alibali MW, Chu MY (2017) How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychol Rev* 124:245–266.
- Lau EF, Phillips C, Poeppel D (2008) A cortical network for semantics: (de)constructing the N400. *Nat Rev Neurosci* 9:920–933.
- Marco DD, De Stefani E, Gentilucci M (2015) Gesture and word analysis: the same or different processes? *Neuroimage* 117:375–385.

- Morrel-Samuels P, Krauss RM (1992) Word familiarity predicts temporal asynchrony of hand gestures and speech. *J Exp Psychol Learn Mem Cogn* 18:615–622.
- Noesselt T, Rieger JW, Schoenfeld MA, Kanowski M, Hinrichs H, Heinze HJ, Driver J (2007) Audiovisual temporal correspondence modulates human multisensory superior temporal sulcus plus primary sensory cortices. *J Neurosci* 27:11431–11441.
- O'Shea J, Muggleton NG, Cowey A, Walsh V (2004) Timing of target discrimination in human frontal eye fields. *J Cogn Neurosci* 16:1060–1067.
- Obermeier C, Gunter TC (2015) Multisensory integration: the case of a time window of gesture-speech integration. *J Cogn Neurosci* 27:292–307.
- Obermeier C, Holle H, Gunter TC (2011) What iconic gesture fragments reveal about gesture-speech integration: when synchrony is lost, memory can help. *J Cogn Neurosci* 23:1648–1663.
- Ozyurek 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.
- Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat Rev Neurosci* 8:976–987.
- Pitcher D, Walsh V, Yovel G, Duchaine B (2007) TMS evidence for the involvement of the right occipital face area in early face processing. *Curr Biol* 17:1568–1573.
- Pitcher D, Garrido L, Walsh V, Duchaine BC (2008) Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *J Neurosci* 28:8929–8933.
- Ralph MA, Jefferies E, Patterson K, Rogers TT (2017) The neural and computational bases of semantic cognition. *Nat Rev Neurosci* 18:42–55.
- Smith AC, Monaghan P, Huettig F (2017) The multimodal nature of spoken word processing in the visual world: testing the predictions of alternative models of multimodal integration. *J Mem Lang* 93:276–303.
- So WC, Yi-Feng AL, Yap DF, Kheng E, Yap JM (2013) Iconic gestures prime words: comparison of priming effects when gestures are presented alone and when they are accompanying speech. *Front Psychol* 4:779.
- Talsma D (2015) Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Front Integr Neurosci* 9:19.
- 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.
- Willems RM, Ozyurek 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.
- Wu YC, Coulson S (2007) Iconic gestures prime related concepts: an ERP study. *Psychon Bull Rev* 14:57–63.
- Xu X, Hanganu-Opatz IL, Bieler M (2020) Cross-talk of low-level sensory and high-level cognitive processing: development, mechanisms, and relevance for cross-modal abilities of the brain. *Front Neurobot* 14:7.
- Yap DF, So WC, Yap JM, Tan YQ, Teoh RL (2011) Iconic gestures prime words. *Cogn Sci* 35:171–183.
- Zhao WY, Riggs K, Schindler I, Holle H (2018) Transcranial magnetic stimulation over left inferior frontal and posterior temporal cortex disrupts gesture-speech integration. *J Neurosci* 38:1891–1900.