

Progressive neural engagement within the IFG-pMTG circuit as gesture and speech entropy and MI advances

Reviewed Preprint

v3 • July 16, 2025

Revised by authors

Reviewed Preprint

v2 • May 19, 2025

Reviewed Preprint

v1 • December 13, 2024

Wanying Zhao , Zhouyi Li, Xiang Li, Yi Du 

State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China • Department of Psychology, University of Chinese Academy of Sciences, Beijing, China • School of Psychology, Central China Normal University, Wuhan, China • Chinese Institute for Brain Research, Beijing, China

 https://en.wikipedia.org/wiki/Open_access
 Copyright information

eLife Assessment

This **useful** study uses brain stimulation and electroencephalography to study speech-gesture integration. It investigates the role of frontotemporal regions in integrating linguistic and extra-linguistic information during communication, focusing on the inferior frontal gyrus and posterior middle temporal gyrus. Reliance on activation patterns of tightly-coupled brain regions over short timescales leads to **incomplete** support for the study's conclusions due to conceptual and methodological limitations.

<https://doi.org/10.7554/eLife.99416.3.sa3>

Abstract

Semantic representation emerges from distributed multisensory modalities, yet a comprehensive understanding of the functional changing pattern within convergence zones or hubs integrating multisensory semantic information remains elusive. In this study, employing information-theoretic metrics, we quantified gesture and speech information, alongside their interaction, utilizing entropy and mutual information (MI). Neural activities were assessed via interruption effects induced by High-Definition transcranial direct current stimulation (HD-tDCS). Additionally, chronometric double-pulse transcranial magnetic stimulation (TMS) and high-temporal event-related potentials were utilized to decipher dynamic neural changes resulting from various information contributors. Results showed gradual inhibition of both inferior frontal gyrus (IFG) and posterior middle temporal gyrus (pMTG) as degree of gesture-speech integration, indexed by MI, increased. Moreover, a time-sensitive and staged progression of neural engagement was observed, evidenced by distinct correlations between neural activity patterns and entropy measures of speech and gesture, as well as MI, across early sensory and lexico-semantic processing stages. These findings illuminate the gradual nature of neural activity during multisensory gesture-speech semantic processing, shaped by dynamic gesture constraints and speech encoding, thereby offering insights into the neural mechanisms underlying multisensory language processing.

Introduction

Semantic representation, distinguished by its cohesive conceptual nature, emerges from distributed modality-specific regions. Consensus acknowledges the presence of ‘convergence zones’ within the temporal and inferior parietal areas¹, or the ‘semantic hub’ located in the anterior temporal lobe², pivotal for integrating, converging, or distilling multimodal inputs. Contemporary theories frame the semantic processing as a dynamic sequence of neural states³, shaped by systems that are finely tuned to the statistical regularities inherent in sensory inputs⁴. These regularities enable the brain to evaluate, weight, and integrate multisensory information, optimizing the reliability of individual sensory signals⁵. However, sensory inputs available to the brain are often incomplete and uncertain, necessitating adaptive neural adjustments to resolve these ambiguities⁶. In this context, neuronal activity is thought to be linked to the probability density of sensory information, with higher levels of uncertainty resulting in the engagement of a broader population of neurons, thereby reflecting the brain’s adaptive capacity to handle diverse possible interpretations^{7,8}. Although the role of ‘convergence zones’ and ‘semantic hubs’ in integrating multimodal inputs is well established, the precise functional patterns of neural activity in response to the distribution of unified multisensory information—along with the influence of unisensory signals—remain poorly understood.

To this end, we developed an analytic approach to directly probe the cortical engagement during multisensory gesture-speech semantic integration. Even though gestures convey information in a global-synthetic way, while speech conveys information in a linear segmented way, there exists a bidirectional semantic influence between the two modalities^{9,10}. Gesture is regarded as ‘part of language’¹¹ or functional equivalents of lexical units that alternate and integrate with speech into a ‘single unification space’ to convey a coherent meaning^{12–14}. Empirical studies have investigated the semantic integration between gesture and speech by manipulating their semantic relationship^{15–18} and revealed a mutual interaction between them^{19–21} as reflected by the N400 latency and amplitude¹⁴ as well as common neural underpinnings in the left inferior frontal gyrus (IFG) and posterior middle temporal gyrus (pMTG)^{15,22,23}.

Building on these insights, the present study quantified the amount of information from both sources and their interaction adopting the information-theoretic complexity metrics of *entropy* and *mutual information* (MI). Unisensory Entropy measures the disorder or randomness of information and serves as an index of the uncertainty in modality-specific representations of gesture or speech activated by an event²⁴. MI assesses shared information between modalities²⁵, indicating multisensory convergence and acting as an index of gesture-speech integration.

To investigate the neural mechanisms underlying gesture-speech integration, we conducted three experiments to assess how neural activity correlates with distributed multisensory integration, quantified using information-theoretic measures of MI. Additionally, we examined the contributions of unisensory signals in this process, quantified through unisensory entropy.

Experiment 1 employed high-definition transcranial direct current stimulation (HD-tDCS) to administer Anodal, Cathodal and Sham stimulation to either the IFG or the pMTG. HD-tDCS induces membrane depolarization with anodal stimulation and membrane hyperpolarization with cathodal stimulation²⁶, thereby increasing or decreasing cortical excitability in the targeted brain area, respectively. This experiment aimed to determine whether the overall facilitation (Anodal-tDCS minus Sham-tDCS) and/or inhibitory (Cathodal-tDCS minus Sham-tDCS) of these integration hubs is modulated by the degree of gesture-speech integration, as measured by MI.

Given the differential involvement of the IFG and pMTG in gesture-speech integration, shaped by top-down gesture predictions and bottom-up speech processing²³, **Experiment 2** was designed to further assess whether the activity of these regions was associated with relevant informational matrices. To this end, we employed chronometric double-pulse transcranial magnetic stimulation, which is known to transiently reduce cortical excitability at the inter-pulse interval²⁷. Within a temporal period broad enough to capture the full duration of gesture-speech integration²⁸, we targeted specific timepoints previously implicated in integrative processing within IFG and pMTG²³. This allowed us to test whether the inhibitory effects of TMS were correlated with unisensory entropy or the multisensory convergence index (MI).

Experiment 3 complemented these investigations by focusing on the temporal dynamics of neural responses during semantic processing, leveraging high-temporal event-related potentials (ERPs). This experiment investigated how distinct information contributors modulated specific ERP components associated with semantic processing. These components included the early sensory effects as P1 and N1–P2^{29,30}, the N400 semantic conflict effect^{14,30,31}, and the late positive component (LPC) reconstruction effect^{32,33}. By integrating these ERP findings with results from Experiments 1 and 2, Experiment 3 aimed to provide a more comprehensive understanding of how gesture-speech integration is modulated by neural dynamics.

Material and methods

Participants

Ninety-eight young Chinese participants signed written informed consent forms and took part in the present study (Experiment 1: 29 females, 23 males, age = 20 ± 3.40 years; Experiment 2: 11 females, 13 males, age = 23 ± 4.88 years; Experiment 3: 12 females, 10 males, age = 21 ± 3.53 years). All of the participants were right-handed (Experiment 1: laterality quotient (LQ)³⁴ = 88.71 ± 13.14 ; Experiment 2: LQ = 89.02 ± 13.25 ; Experiment 3: LQ = 88.49 ± 12.65), had normal or corrected-to-normal vision and were paid ¥100 per hour for their participation. All experiments were approved by the Ethics Committee of the Institute of Psychology, Chinese Academy of Sciences.

Stimuli

Twenty gestures (**Appendix Table 1**) with 20 semantically congruent speech signals taken from previous study²³ were used. The stimuli set were recorded from two native Chinese speakers (1 male, 1 female). To validate the stimuli, 30 participants were recruited to replicate the multisensory index of semantic congruency effect, hypothesizing that reaction times for semantically incongruent gesture-speech pairs would be significantly longer than those for congruent pairs. The results confirmed this hypothesis, with a significantly ($t(29) = 7.16$, $p < 0.001$) larger reaction time when participants were asked to judge the gender of the speaker if gesture contained incongruent semantic information with speech (a ‘cut’ gesture paired with speech word ‘喷 pen1 (spray)’; mean = 554.51 ms, SE = 11.65) relative to when they were semantically congruent (a ‘cut’ gesture paired with ‘剪 jian3 (cut)’ word: mean = 533.90 ms, SE = 12.02)²³.

Additionally, two separate pre-tests with 30 subjects in each (pre-test 1: 16 females, 14 males, age = 24 ± 4.37 years; pre-test 2: 15 females, 15 males, age = 22 ± 3.26 years) were conducted to determine the comprehensive values of gesture and speech. Participants were presented with segments of increasing duration, beginning at 40 ms, and were prompted to provide a single verb to describe either the isolated gesture they observed (pre-test 1) or the isolated speech they heard (pre-test 2). For each gesture or speech, the action verb consistently provided by participants across four to six consecutive repetitions—with the number of repetitions varied to mitigate

learning effects—was considered the comprehensive response for the gesture or speech. The initial instance duration was marked as the discrimination point (DP) for gesture (mean = 183.78 ± 84.82ms) or the identification point (IP) for speech (mean = 176.40 ± 66.21ms) (**Figure 1A top** [↗](#)).

To quantify information content, comprehensive responses for each item were converted into Shannon's entropy (H) as a measure of information richness (**Figure 1A bottom** [↗](#)). With no significant gender differences observed in both gesture ($t(20) = 0.21, p = 0.84$) and speech ($t(20) = 0.52, p = 0.61$), responses were aggregated across genders, resulting in 60 answers per item (**Appendix Table 2** [↗](#)). Here, $p(x_i)$ and $p(y_i)$ represent the distribution of 60 answers for a given gesture (**Appendix Table 2B** [↗](#)) and speech (**Appendix Table 2A** [↗](#)), respectively. High entropy indicates diverse answers, reflecting broad representation, while low entropy suggests focused lexical recognition for a specific item (**Figure 2B** [↗](#)). MI was used to measure the overlap between gesture and speech information, calculated by subtracting the entropy of the combined gesture-speech dataset (Entropy(gesture + speech)) from the sum of their individual entropies (Entropy(gesture) + Entropy(speech)) (see **Appendix Table 2C** [↗](#)). For specific gesture-speech combinations, equivalence between the combined entropy and the sum of individual entropies (gesture or speech) indicates absence of overlap in response sets. Conversely, significant overlap, denoted by a considerable number of shared responses between gesture and speech datasets, leads to a noticeable discrepancy between combined entropy and the sum of gesture and speech entropies. Elevated MI values thus signify substantial overlap, indicative of a robust mutual interaction between gesture and speech.

Additionally, the number of responses provided for each gesture and speech, as well as the total number of combined responses, were also recorded. The quantitative data for each stimulus, including gesture entropy, speech entropy, joint entropy, MI, and the respective counts, are presented in **Appendix Table 3** [↗](#).

To determine whether entropy or MI values corresponds to distinct neural changes, the current study first aggregated neural responses (including inhibition effects of tDCS and TMS or ERP amplitudes) that shared identical entropy or MI values, prior to conducting correlational analyses.

Experimental procedure

Given that gestures induce a semantic priming effect on concurrent speech³⁵ [↗](#), this study utilized a semantic priming paradigm in which speech onset was aligned with the DP of each gesture²³ [↗](#),³⁵ [↗](#), the point at which the gesture transitions into a lexical form³⁶ [↗](#). The gesture itself began at the stroke phase, a critical moment when the gesture conveys its primary semantic content³⁶ [↗](#).

An irrelevant factor of gender congruency (e.g., a man making a gesture combined with a female voice) was created²² [↗](#),²³ [↗](#),³⁷ [↗](#). This involved aligning the gender of the voice with the corresponding gender of the gesture in either a congruent (e.g., male voice paired with a male gesture) or incongruent (e.g., male voice paired with a female gesture) manner. This approach served as a direct control mechanism, facilitating the investigation of the automatic and implicit semantic interplay between gesture and speech³⁷ [↗](#). In light of previous findings indicating a distinct TMS-disruption effect on the semantic congruency of gesture-speech interactions²³ [↗](#), both semantically congruent and incongruent pairs were included in Experiment 1 and Experiment 2. Experiment 3, conversely, exclusively utilized semantically congruent pairs to elucidate ERP metrics indicative of nuanced semantic progression.

Gesture-speech pairs were presented randomly using Presentation software (www.neurobs.com [↗](#)). Participants were asked to look at the screen but respond with both hands as quickly and accurately as possible merely to the gender of the voice they heard. The RT and the

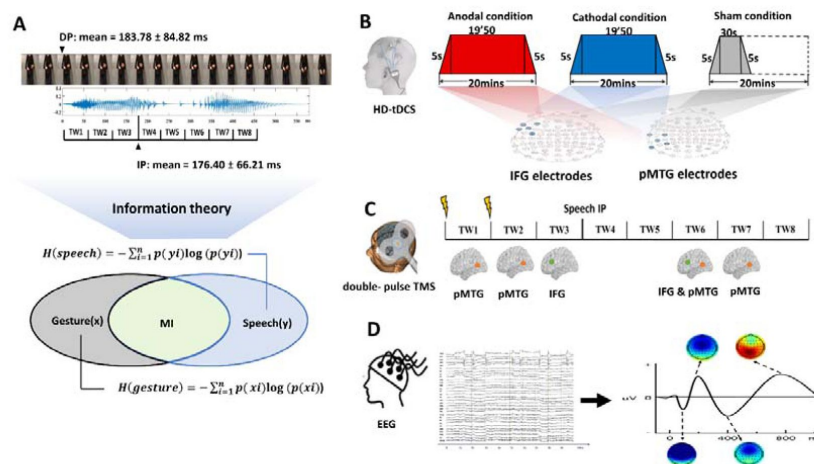


Figure 1.

Experimental design, and stimulus characteristics.

(A) Experimental stimuli. Twenty gestures were paired with 20 relevant speech stimuli. Two separate Pre-tests were executed to define the minimal length of each gesture and speech required for semantic identification, namely, the discrimination point (DP) of gesture and the identification point (IP) of speech. Overall, a mean of 183.78 ms (SD = 84.82) was found for the DP of gestures and the IP of speech was 176.40 ms (SD = 66.21). The onset of speech was set at the gesture DP. Responses for each item were assessed utilizing information-theoretic complexity metrics to quantify the information content of both gesture and speech during integration, employing entropy and MI.

Procedure(B) of Experiment 1. HD-tDCS, including Anodal, Cathodal, or Sham conditions, was administered to the IFG or pMTG) using a 4 * 1 ring-based electrode montage. Electrode F7 targeted the IFG, with return electrodes placed on AF7, FC5, F9, and FT9. For pMTG stimulation, TP7 was targeted, with return electrodes positioned on C5, P5, T9, and P9. Sessions lasted 20 minutes, with a 5-second fade-in and fade-out, while the Sham condition involved only 30 seconds of stimulation.

(C) Procedure of Experiment 2. Eight time windows (TWs, duration = 40 ms) were segmented in relative to the speech IP. Among the eight TWs, five (TW1, TW2, TW3, TW6, and TW7) were chosen based on the significant results in our prior study²³. Double-pulse TMS was delivered over each of the TW of either the pMTG or the IFG.

(D) Procedure of Experiment 3. Semantically congruent gesture-speech pairs were presented randomly with Electroencephalogram (EEG) recorded simultaneously. Epochs were time locked to the onset of speech and lasted for 1000 ms. A 200 ms pre-stimulus baseline correction was applied before the onset of gesture stroke. Various elicited components were hypothesized.

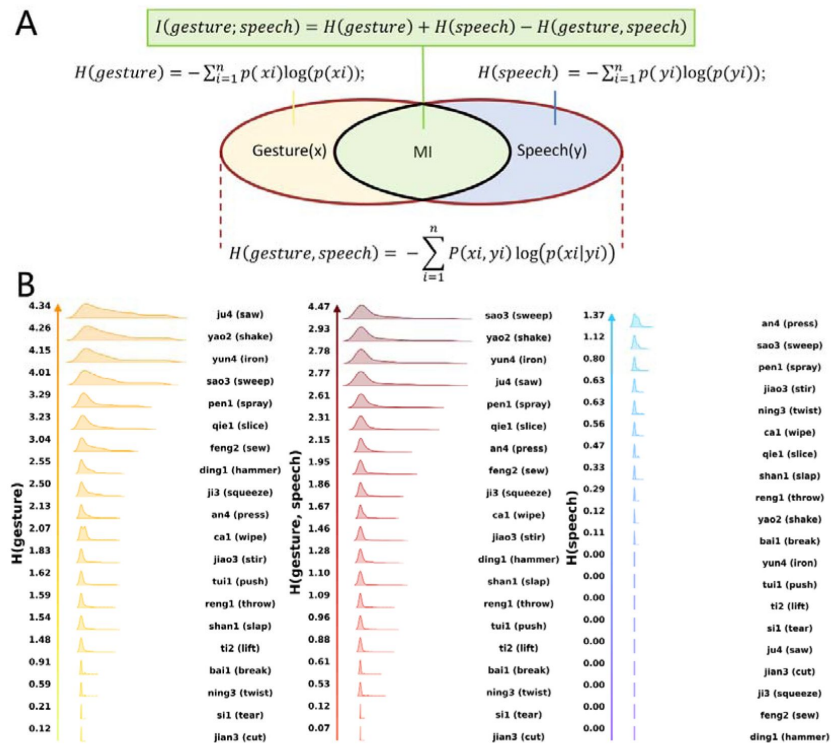


Figure 2.

Quantification formulas (A) and distributions of each stimulus in Shannon's entropy (B).

Two separate pre-tests ($N = 30$) were conducted to assign a single verb for describing each of the isolated 20 gestures and 20 speech items. Responses provided for each item were transformed into Shannon's entropy using a relative quantification formula. Gesture (**B left**) and speech (**B right**) entropy quantify the randomness of gestural or speech information, representing the uncertainty of probabilistic representation activated when a specific stimulus occurs. Joint entropy (**B middle**) captures the widespread nature of the two sources of information combined. Mutual information (MI) was calculated as the difference between joint entropy with gesture entropy and speech entropy combined (**A**), thereby capturing the overlap of gesture and speech and representing semantic integration.

button being pressed were recorded. The experiment started with a fixation cross presented on the center of the screen, which lasted for 0.5-1.5 sec.

Experiment 1: HD-tDCS protocol and data analysis

Participants were divided into two groups, with each group undergoing HD-tDCS stimulation at different target sites (IFG or pMTG). Each participant completed three experimental sessions, spaced one week apart, during which 480 gesture-speech pairs were presented across various conditions. In each session, participants received one of three types of HD-tDCS stimulation: Anodal, Cathodal, or Sham. The order of stimulation site and type was counterbalanced using a Latin square design to control for potential order effects.

HD-tDCS protocol employed a constant current stimulator (The Starstim 8 system) delivering stimulation at an intensity of 2000mA. A 4 * 1 ring-based electrode montage was utilized, comprising a central electrode (stimulation) positioned directly over the target cortical area and four return electrodes encircling it to provide focused stimulation. Building on a meta-analysis of prior fMRI studies examining gesture-speech integration²², we targeted Montreal Neurological Institute (MNI) coordinates for the left IFG at (-62, 16, 22) and the pMTG at (-50, -56, 10). In the stimulation protocol for HD-tDCS, the IFG was targeted using electrode F7 as the optimal cortical projection site³⁸, with four return electrodes placed at AF7, FC5, F9, and FT9. For the pMTG, TP7 was selected as the cortical projection site³⁸, with return electrodes positioned at C5, P5, T9, and P9. The stimulation parameters included a 20-minute duration with a 5-second fade-in and fade-out for both Anodal and Cathodal conditions. The Sham condition involved a 5-second fade-in followed by only 30 seconds of stimulation, then 19'20 minutes of no stimulation, and finally a 5-second fade-out (**Figure 1B**). Stimulation was controlled using NIC software, with participants blinded to the stimulation conditions.

All incorrect responses (702 out of the total number of 24960, 2.81% of trials) were excluded. To eliminate the influence of outliers, a 2SD trimmed mean for every participant in each session was also calculated. To examine the relationship between the degree of information and neural responses, we conducted Pearson correlation analyses using a sample of 20 sets. Neural responses were quantified based on the effects of HD-tDCS (active tDCS minus sham tDCS) on the semantic congruency effect, defined as the difference in reaction times between semantic incongruent and congruent conditions ($Rt(\text{incongruent}) - Rt(\text{congruent})$). This effect served as an index of multisensory integration³⁷ within the left IFG and pMTG. The variation in information was assessed using three information-theoretic metrics. To account for potential confounds related to multiple candidate representations, we conducted partial correlation analyses between the tDCS effects and gesture entropy, speech entropy, and MI, controlling for the number of responses provided for each gesture and speech, as well as the total number of combined responses. Given that HD-tDCS induces overall disruption at the targeted brain regions, we hypothesized that the neural activity within the left IFG and pMTG would be progressively affected by varying levels of multisensory convergence, as indexed by MI. Moreover, we hypothesized that the modulation of neural activity by MI would differ between the left IFG and pMTG, as reflected in the differential modulation of response numbers in the partial correlations, highlighting their distinct roles in semantic processing³⁹. False discovery rate (FDR) correction was applied for multiple comparisons.

Experiment 2: TMS protocol and data analysis

Experiment 2 involved 800 gesture-speech pairs, presented across 15 blocks over three days, with one week between sessions. Stimulation was administered at three different sites (IFG, pMTG, or Vertex). Within the time windows (TWs) spanning the gesture-speech integration period, five TWs that exhibited selective disruption of integration were selected: TW1 (-120 to -80 ms relative to the

speech identification point), TW2 (-80 to -40 ms), TW3 (-40 to 0 ms), TW6 (80 to 120 ms), and TW7 (120 to 160 ms)²³ (Figure 1C). The order of stimulation site and TW was counterbalanced using a Latin square design.

At an intensity of 50% of the maximum stimulator output, double-pulse TMS was delivered via a 70 mm figure-eight coil using a Magstim Rapid² stimulator (Magstim, UK). High-resolution (1 × 1 × 0.6 mm) T1-weighted MRI scans were obtained using a Siemens 3T Trio/Tim Scanner for image-guided TMS navigation. Frameless stereotaxic procedures (BrainSight 2; Rogue Research) allowed real-time stimulation monitoring. To ensure precision, individual anatomical images were manually registered by identifying the anterior and posterior commissures. Subject-specific target regions were defined using trajectory markers in the MNI coordinate system. Vertex was used as control.

All incorrect responses (922 out of the total number of 19200, 4.8% of trials) were excluded. We focused our analysis on Pearson correlations of the TMS interruption effects (active TMS minus vertex TMS) of the semantic congruency effect with the gesture entropy, speech entropy or MI. To control for potential confounds, partial correlations were also performed between the TMS effects and gesture entropy, speech entropy, and MI, controlling for the number of responses for each gesture and speech, as well as the total number of combined responses. By doing this, we can determine how the time-sensitive contribution of the left IFG and pMTG to gesture–speech integration was affected by gesture and speech information distribution. FDR correction was applied for multiple comparisons.

Experiment 3: Electroencephalogram (EEG) recording and data analysis

Experiment 3, comprising a total of 1760 gesture-speech pairs, was completed in a single-day session. EEG were recorded from 48 Ag/AgCl electrodes mounted in a cap according to the 10-20 system⁴⁰, amplified with a PORTI-32/MREFA amplifier (TMS International B.V., Enschede, NL) and digitized online at 500 Hz (bandpass, 0.01-70 Hz). EEGLAB, a MATLAB toolbox, was used to analyze the EEG data⁴¹. Vertical and horizontal eye movements were measured with 4 electrodes placed above the left eyebrow, below the left orbital ridge and at bilateral external canthus. All electrodes were referenced online to the left mastoid. Electrode impedance was maintained below 5 KΩ. The average of the left and right mastoids was used for re-referencing. A high-pass filter with a cutoff of 0.05 Hz and a low-pass filter with a cutoff of 30 Hz were applied. Semi-automated artifact removal, including independent component analysis (ICA) for identifying components of eye blinks and muscle activity, was performed (Figure 1D). Participants with rejected trials exceeding 30% of their total were excluded from further analysis.

All incorrect responses were excluded (147 out of 1760, 8.35% of trials). To eliminate the influence of outliers, a 2 SD trimmed mean was calculated for every participant in each condition. Data were epoched from the onset of speech and lasted for 1000 ms. To ensure a clean baseline with no stimulus presented, a 200 ms pre-stimulus baseline correction was applied before gesture onset.

To consolidate the data, we conducted both a traditional region-of-interest (ROI) analysis, with ROIs defined based on a well-established work⁴², and a cluster-based permutation approach, which utilizes data-driven permutations to enhance robustness and address multiple comparisons.

For the traditional ROI analysis, grand-average ERPs at electrode Cz were compared between the higher (≥50%) and lower (<50%) halves for gesture entropy (Figure 5A1), speech entropy (Figure 5B1), and MI (Figure 5C1). Consequently, four ERP components were determined: the P1 effect observed within the time window of 0-100 ms^{29,30}, the N1-P2 effect observed between 150-250ms^{29,30}, the N400 within the interval of 250-450ms^{14,30,31}, and the LPC spanning from 550-1000ms^{32,33}. Additionally, seven regions-of-interest (ROIs) were defined in

order to locate the modulation effect on each ERP component: left anterior (LA): F1, F3, F5, FC1, FC3, and FC5; left central (LC): C1, C3, C5, CP1, CP3, and CP5; left posterior (LP): P1, P3, P5, PO3, PO5, and O1; right anterior (RA): F2, F4, F6, FC2, FC4, and FC6; right central (RC): C2, C4, C6, CP2, CP4, and CP6; right posterior (RP): P2, P4, P6, PO4, PO6, and O2; and midline electrodes (ML): Fz, FCz, Cz, Pz, Oz, and CPz⁴².

Subsequently, cluster-based permutation tests⁴³ in Fieldtrip was further used to determine the significant clusters of adjacent time points and electrodes of ERP amplitude between the higher and lower halves of gesture entropy, speech entropy and MI, respectively. The electrode-level type I error threshold was set to 0.025. Cluster-level statistic was estimated through 5000 Monte Carlo simulations, where the cluster-level statistic is the sum of T-values for each stimulus within a cluster. The cluster-level type I error threshold was set to 0.05. Clusters with a p-value less than the critical alpha-level are considered to be conditionally different.

Paired t-tests were conducted to compare the lower and upper halves of each information model for the averaged amplitude within each ROI or cluster across the four ERP time windows, separately. Pearson correlations were computed between each model value and the averaged ERP amplitudes in each ROI or cluster. Additionally, partial correlations were conducted, accounting for the number of responses for each respective metric. FDR correction was applied for multiple comparisons.

Results

Experiment 1: Modulation of left pMTG and IFG engagement by gradual changes in gesture-speech semantic information

In the IFG, one-way ANOVA examining the effects of three tDCS conditions (Anodal, Cathodal, or Sham) on semantic congruency (RT (semantic incongruent) – RT (semantic congruent)) demonstrated a significant main effect of stimulation condition ($F(2, 75) = 3.673$, $p = 0.030$, $\eta^2 = 0.089$). Post hoc paired t-tests indicated a significantly reduced semantic congruency effect between the Cathodal condition and the Sham condition ($t(26) = -3.296$, $p = 0.003$, 95% CI = [-11.488, 4.896]) (Figure 3A left). Subsequent Pearson correlation analysis revealed that the reduced semantic congruency effect was progressively associated with the MI, evidenced by a significant correlation between the Cathodal-tDCS effect (Cathodal-tDCS minus Sham- tDCS) and MI ($r = -0.595$, $p = 0.007$, 95% CI = [-0.995, -0.195]) (Figure 3B). Additionally, a similar correlation was observed between the Cathodal-tDCS effect and the total response number ($r = -0.543$, $p = 0.016$, 95% CI = [-0.961, -0.125]).

However, partial correlation analysis, controlling for the total response number, revealed that the initially significant correlation between the Cathodal-tDCS effect and MI was no longer significant ($r = -0.303$, $p = 0.222$, 95% CI = [-0.770, 0.164]). This suggests that the observed relationship between Cathodal-tDCS and MI may be confounded by semantic control difficulty, as reflected by the total number of responses. Specifically, the reduced activity in the IFG under Cathodal-tDCS may be driven by variations in the difficulty of semantic control rather than a direct modulation of MI.

In the pMTG, a one-way ANOVA assessing the effects of three tDCS conditions on semantic congruency also revealed a significant main effect of stimulation condition ($F(2, 75) = 3.250$, $p = 0.044$, $\eta^2 = 0.080$). Subsequent paired t-tests identified a significantly reduced semantic congruency effect between the Cathodal condition and the Sham condition ($t(25) = -2.740$, $p = 0.011$, 95% CI = [-11.915, 6.435]) (Figure 3A right). Moreover, a significant correlation was observed between the Cathodal-tDCS effect and MI ($r = -0.457$, $p = 0.049$, 95% CI = [-0.900, -0.014]) (Figure 3B).

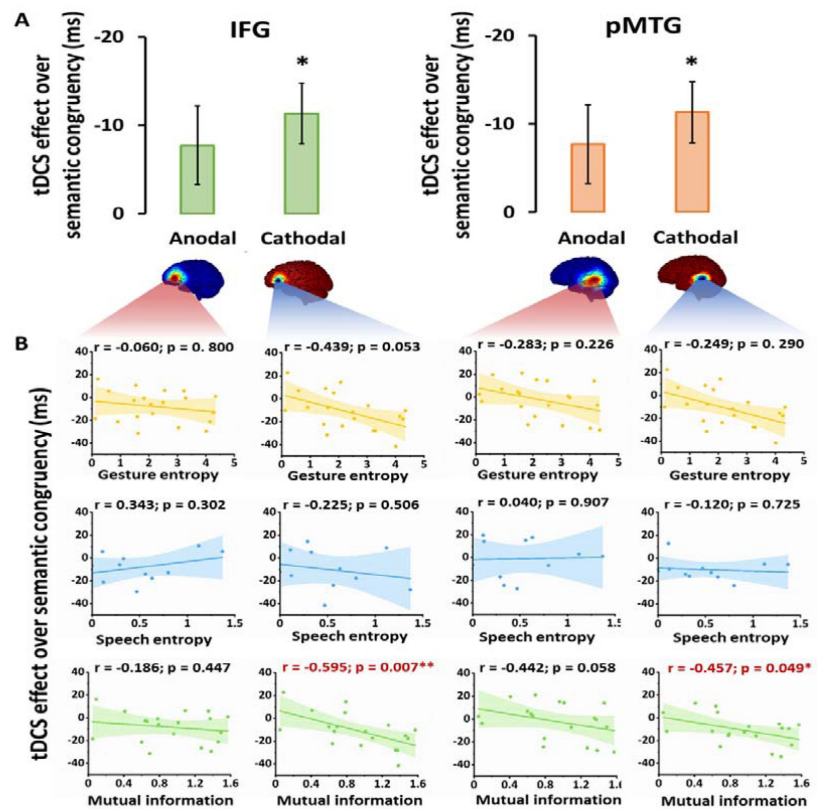


Figure 3.

tDCS effect over semantic congruency.

(A) tDCS effect was defined as active-tDCS minus sham-tDCS. The semantic congruency effect was calculated as the reaction time (RT) difference between semantically incongruent and semantically congruent pairs ($Rt(\text{incongruent}) - Rt(\text{congruent})$).

(B) Correlations of the tDCS effect over the semantic congruency effect with three information models (gesture entropy, speech entropy and MI) are displayed with best-fitting regression lines. Significant correlations are marked in red. * $p < 0.05$, ** $p < 0.01$ after FDR correction.

Importantly, the reduced activity in the pMTG under Cathodal-tDCS was not influenced by the total response number, as indicated by the non-significant correlation ($r = -0.253$, $p = 0.295$, 95% CI = [-0.735, 0.229]). This finding was further corroborated by the unchanged significance in the partial correlation between Cathodal-tDCS and MI, when controlling for the total response number ($r = -0.472$, $p = 0.048$, 95% CI = [-0.903, -0.041]).

RTs of congruent and incongruent trials of IFG and pMTG in each of the stimulation conditions were shown in **Appendix Table 4A** [↗](#).

Experiment 2: Time-sensitive modulation of left pMTG and IFG engagements by gradual changes in gesture-speech semantic information

A 2 (TMS effect: active - Vertex) $\times$ 5 (TW) ANOVA on semantic congruency revealed a significant interaction between TMS effect and TW ($F(3.589, 82.538) = 3.273$, $p = 0.019$, $\eta^2 = 0.125$). Further t -tests identified a significant TMS effect over the pMTG in TW1 ($t(23) = -3.068$, $p = 0.005$, 95% CI = [-6.838, 0.702]), TW2 ($t(23) = -2.923$, $p = 0.008$, 95% CI = [-6.490, 0.644]), and TW7 ($t(23) = -2.005$, $p = 0.047$, 95% CI = [-5.628, 1.618]). In contrast, a significant TMS effect over the IFG was found in TW3 ($t(23) = -2.335$, $p = 0.029$, 95% CI = [-5.928, 1.258]), and TW6 ($t(23) = -4.839$, $p < 0.001$, 95% CI = [-7.617, -2.061]) (**Figure 4A** [↗](#)). Raw RTs of congruent and incongruent trials were shown in **Appendix Table 4B** [↗](#).

Additionally, a significant negative correlation was found between the TMS effect (a larger negative TMS effect signifies a greater disruption of the integration process) and speech entropy when the pMTG was inhibited in TW2 ($r = -0.792$, $p = 0.004$, 95% CI = [-1.252, -0.331]). Meanwhile, when the IFG activity was interrupted in TW6, a significant negative correlation was found between the TMS effect and gesture entropy ($r = -0.539$, $p = 0.014$, 95% CI = [-0.956, -0.122]), speech entropy ($r = -0.664$, $p = 0.026$, 95% CI = [-1.255, -0.073]), and MI ($r = -0.677$, $p = 0.001$, 95% CI = [-1.054, -0.300]) (**Figure 4B** [↗](#)).

Notably, inhibition of pMTG activity in TW2 was not influenced by the number of speech responses ($r = -0.539$, $p = 0.087$, 95% CI = [-1.145, 0.067]). However, the number of speech responses did affect the modulation of speech entropy on the pMTG inhibition effect in TW2. This was evidenced by the non-significant partial correlation between pMTG inhibition and speech entropy when controlling for speech response number ($r = -0.218$, $p = 0.545$, 95% CI = [-0.563, 0.127]).

In contrast, the interrupted IFG activity in TW6 appeared to be consistently influenced by the confound of semantic control difficulty. This was reflected in the significant correlation with both gesture response number ($r = -0.480$, $p = 0.032$, 95% CI = [-0.904, -0.056]), speech response number ($r = -0.729$, $p = 0.011$, 95% CI = [-1.221, -0.237]), and total response number ($r = -0.591$, $p = 0.008$, 95% CI = [-0.993, -0.189]). Additionally, partial correlation analyses revealed non-significant relationship between interrupted IFG activity in TW6 and gesture entropy ($r = -0.369$, $p = 0.120$, 95% CI = [-0.810, -0.072]), speech entropy ($r = -0.455$, $p = 0.187$, 95% CI = [-1.072, 0.162]), and MI ($r = -0.410$, $p = 0.091$, 95% CI = [-0.856, -0.036]) when controlling for response numbers.

Experiment 3: Temporal modulation of P1, N1-P2, N400 and LPC components by gradual changes in gesture-speech semantic information

Topographical maps illustrating amplitude differences between the lower and higher halves of speech entropy demonstrate a central-posterior P1 amplitude (0-100 ms, **Figure 5B** [↗](#)). Aligning with prior findings²⁹ [↗](#), the paired t -tests demonstrated a significantly larger P1 amplitude within the ML ROI ($t(22) = 2.510$, $p = 0.020$, 95% confidence interval (CI) = [1.66, 3.36]) when contrasting

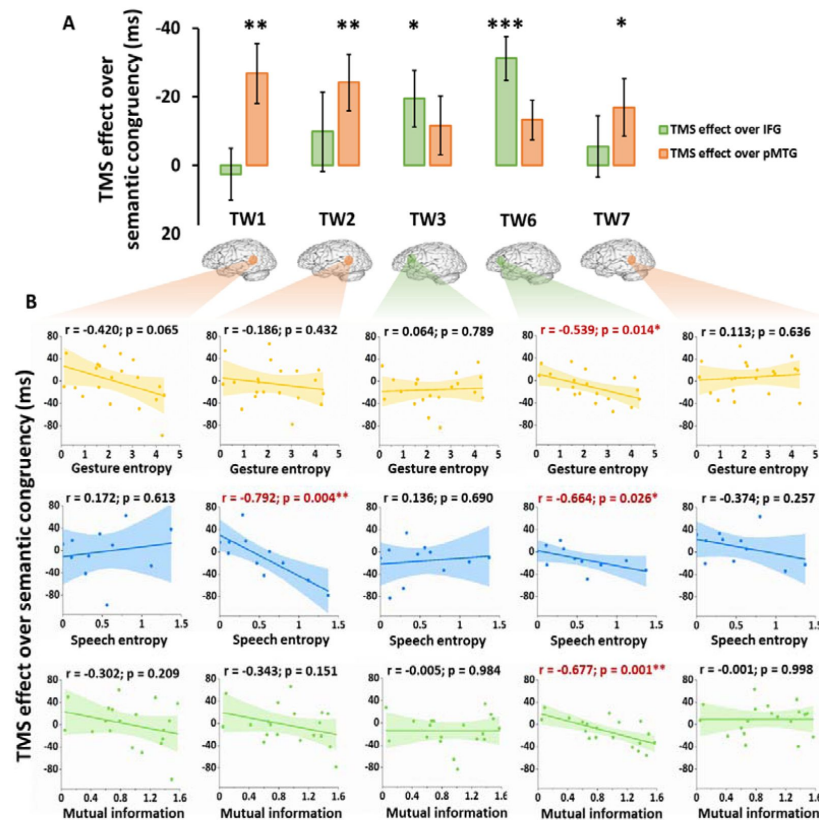


Figure 4.

TMS impacts on semantic congruency effect across various time windows (TW).

(A) Five time windows (TWs) showing selective disruption of gesture-speech integration were chosen: TW1 (-120 to -80 ms relative to speech identification point), TW2 (-80 to -40 ms), TW3 (-40 to 0 ms), TW6 (80 to 120 ms), and TW7 (120 to 160 ms). TMS effect was defined as active-TMS minus vertex-TMS. The semantic congruency effect was calculated as the reaction time (RT) difference between semantically incongruent and semantically congruent pairs.

(B) Correlations of the TMS effect over the semantic congruency effect with three information models (gesture entropy, speech entropy and MI) are displayed with best-fitting regression lines. Significant correlations are marked in red. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ after FDR correction.

stimuli with higher 50% speech entropy against those with lower 50% speech entropy (**Figure 5D1 left**). Subsequent correlation analyses unveiled a significant increase in the P1 amplitude with the rise in speech entropy within the ML ROI ($r = 0.609$, $p = 0.047$, 95% CI = [0.039, 1.179], **Figure 5D1 right**). Furthermore, a cluster of neighboring time-electrode samples exhibited a significant contrast between the lower 50% and higher 50% of speech entropy, revealing a P1 effect spanning 16 to 78 ms at specific electrodes (FC2, FCz, C1, C2, Cz, and CPz, **Figure 5D2 middle**) ($t(22) = 2.754$, $p = 0.004$, 95% confidence interval (CI) = [1.65, 3.86], **Figure 5D2 left**), with a significant correlation with speech entropy ($r = 0.636$, $p = 0.035$, 95% CI = [0.081, 1.191], **Figure 5D2 right**).

Additionally, topographical maps comparing the lower 50% and higher 50% gesture entropy revealed a frontal N1-P2 amplitude (150-250 ms, **Figure 5A**). In accordance with previous findings on bilateral frontal N1-P2 amplitude²⁹, paired t-tests displayed a significantly larger amplitude for stimuli with lower 50% gesture entropy than with higher 50% entropy in both ROIs of LA ($t(22) = 2.820$, $p = 0.011$, 95% CI = [2.21, 3.43]) and RA ($t(22) = 2.223$, $p = 0.038$, 95% CI = [1.56, 2.89]) (**Figure 5E1 left**). Moreover, a negative correlation was found between N1-P2 amplitude and gesture entropy in both ROIs of LA ($r = -0.465$, $p = 0.039$, 95% CI = [-0.87, -0.06]) and RA ($r = -0.465$, $p = 0.039$, 95% CI = [-0.88, -0.05]) (**Figure 5E1 right**). Additionally, through a cluster-permutation test, the N1-P2 effect was identified between 184 to 202 ms at electrodes FC4, FC6, C2, C4, C6, and CP4 (**Figure 5E2 middle**) ($t(22) = 2.638$, $p = 0.015$, 95% CI = [1.79, 3.48], (**Figure 5E2 left**)), exhibiting a significant correlation with gesture entropy ($r = -0.485$, $p = 0.030$, 95% CI = [-0.91, -0.06], **Figure 5E2 right**).

Furthermore, in line with prior research⁴⁴, a left-frontal N400 amplitude (250-450 ms) was discerned from topographical maps of gesture entropy (**Figure 5A**). Specifically, stimuli with lower 50% values of gesture entropy elicited a larger N400 amplitude in the LA ROI compared to those with higher 50% values ($t(22) = 2.455$, $p = 0.023$, 95% CI = [1.95, 2.96], **Figure 5F1 left**). Concurrently, a negative correlation was noted between the N400 amplitude and gesture entropy ($r = -0.480$, $p = 0.032$, 95% CI = [-0.94, -0.03], **Figure 5F1 right**) within the LA ROI. The identified clusters showing the N400 effect for gesture entropy (282–318 ms at electrodes FC1, FCz, C1, and Cz, **Figure 5F2 middle**) ($t(22) = 2.828$, $p = 0.010$, 95% CI = [2.02, 3.64], **Figure 5F2 left**) also exhibited significant correlation between the N400 amplitude and gesture entropy ($r = -0.445$, $p = 0.049$, 95% CI = [-0.88, -0.01], **Figure 5F2 right**).

Similarly, a left-frontal N400 amplitude (250-450 ms)⁴⁴ was discerned from topographical maps for MI (**Figure 5C**). A larger N400 amplitude in the LA ROI was observed for stimuli with lower 50% values of MI compared to those with higher 50% values ($t(22) = 3.00$, $p = 0.007$, 95% CI = [2.54, 3.46], **Figure 5G1 left**). This was accompanied by a significant negative correlation between N400 amplitude and MI ($r = -0.504$, $p = 0.028$, 95% CI = [-0.97, -0.04], **Figure 5G1 right**) within the LA ROI. The N400 effect for MI, observed in the 294–306 ms window at electrodes F1, F3, Fz, FC1, FC3, FCz, and C1 (**Figure 5G2 middle**) ($t(22) = 2.461$, $p = 0.023$, 95% CI = [1.62, 3.30], **Figure 5G2 left**), also showed a significant negative correlation with MI ($r = -0.569$, $p = 0.011$, 95% CI = [-0.98, -0.16], **Figure 5G2 right**).

Finally, consistent with previous findings³², an anterior LPC effect (550-1000 ms) was observed in topographical maps comparing stimuli with lower and higher 50% speech entropy (**Figure 5B**). The reduced LPC amplitude was evident in the paired t-tests conducted in ROIs of LA ($t(22) = 2.614$, $p = 0.016$, 95% CI = [1.88, 3.35]); LC ($t(22) = 2.592$, $p = 0.017$, 95% CI = [1.83, 3.35]); RA ($t(22) = 2.520$, $p = 0.020$, 95% CI = [1.84, 3.24]); and ML ($t(22) = 2.267$, $p = 0.034$, 95% CI = [1.44, 3.10]) (**Figure 5H1 left**). Simultaneously, a marked negative correlation with speech entropy was evidenced in ROIs of LA ($r = -0.836$, $p = 0.001$, 95% CI = [-1.26, -0.42]); LC ($r = -0.762$, $p = 0.006$, 95% CI = [-1.23, -0.30]); RA ($r = -0.774$, $p = 0.005$, 95% CI = [-1.23, -0.32]) and ML ($r = -0.730$, $p = 0.011$, 95% CI = [-1.22, -0.24]) (**Figure 5H1 right**). Additionally, a cluster with the LPC effect (644–688 ms at electrodes

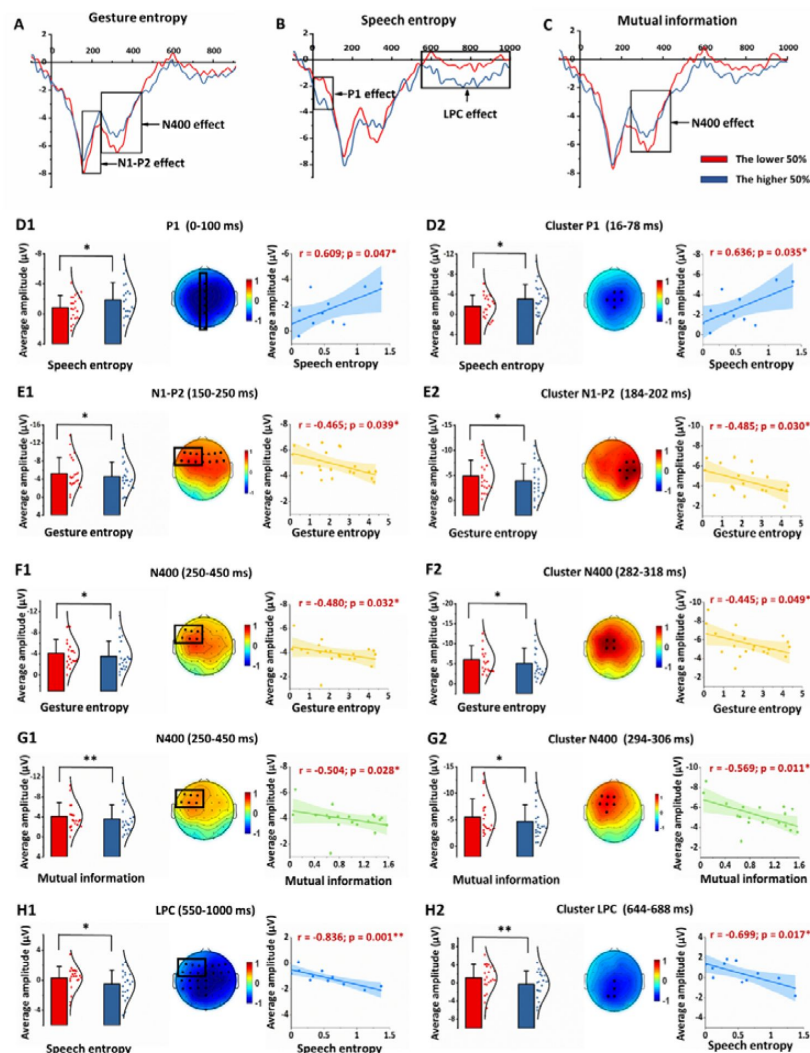


Figure 5.

ERP results of gesture entropy (A), speech entropy (B) or MI (C).

Four ERP components were identified from grand-average ERPs at the Cz electrode, contrasting trials with the lower 50% (red lines) and the higher 50% (blue lines) of gesture entropy, speech entropy or MI. Clusters of adjacent time points and electrodes were subsequently identified within each component using a cluster-based permutation test. Topographical maps depict amplitude differences between the lower and higher halves of each information model, with significant ROIs (**D1-H1 middle**) or electrode clusters (**D2-H2 middle**) highlighted in black. Solid rectangles delineating the ROIs that exhibited the maximal correlation and paired t-values (**D1-H1 middle**). T-test comparisons with normal distribution lines (**left**) and correlations with best-fitting regression lines (**right**) are calculated and illustrated between the average ERP amplitude within the rectangular ROI (**D1-H1**) or the elicited clusters (**D2-H2**) and the three information models individually. * $p < 0.05$, ** $p < 0.01$ after FDR correction.

Cz, CPz, P1, and Pz, **Figure 5H2 middle** ([↗](#)) ($t(22) = 2.754, p = 0.012, 95\% \text{ CI} = [1.50, 4.01]$, **Figure 5H2 left** ([↗](#)) displayed a significant correlation with speech entropy ($r = -0.699, p = 0.017, 95\% \text{ CI} = [-1.24, -0.16]$, **Figure 5H2 right** ([↗](#)).

To clarify potential confounds of semantic control difficulty, partial correlation analyses were conducted to examine the relationship between the elicited ERP components and the relevant information matrices, controlling for response numbers. Results consistently indicated modulation by response numbers in the relationship of ERP components with the information matrix, as evidenced by the non-significant partial correlations between the P1 amplitude (P1 component over ML: $r = -0.574, p = 0.082, 95\% \text{ CI} = [-1.141, -0.007]$) and the P1 cluster ($r = -0.503, p = 0.138, 95\% \text{ CI} = [-1.102, 0.096]$) with speech entropy; the N1-P2 amplitude (N1-P2 component over LA: $r = -0.080, p = 0.746, 95\% \text{ CI} = [-0.554, 0.394]$) and N1-P2 cluster ($r = -0.179, p = 0.464, 95\% \text{ CI} = [-0.647, 0.289]$) with gesture entropy; the N400 amplitude (N400 component over LA: $r = 0.264, p = 0.247, 95\% \text{ CI} = [-0.195, 0.723]$) and N400 cluster ($r = 0.394, p = 0.095, 95\% \text{ CI} = [-0.043, 0.831]$) with gesture entropy; the N400 amplitude (N400 component over LA: $r = -0.134, p = 0.595, 95\% \text{ CI} = [-0.620, 0.352]$) and N400 cluster ($r = -0.034, p = 0.894, 95\% \text{ CI} = [-0.524, 0.456]$) with MI; and the LPC amplitude (LPC component over LA: $r = -0.428, p = 0.217, 95\% \text{ CI} = [-1.054, 0.198]$) and LPC cluster ($r = -0.202, p = 0.575, 95\% \text{ CI} = [-0.881, 0.477]$) with speech entropy.

Discussion

Through mathematical quantification of gesture and speech information using entropy and mutual information (MI), we examined the functional pattern and dynamic neural structure underlying multisensory semantic integration. Our results, for the first time, revealed that the inhibition effect of cathodal-tDCS on the pMTG and IFG correlated with the degree of gesture-speech multisensory convergence, as indexed by MI (**Experiment 1**). Moreover, the gradual neural engagement was found to be time-sensitive and staged, as evidenced by the selectively interrupted time windows (**Experiment 2**) and the distinct correlated ERP components (**Experiment 3**), which were modulated by different information contributors, including unisensory entropy or multisensory MI. These findings significantly expand our understanding of the cortical foundations of statistically regularized multisensory semantic information.

It is widely acknowledged that a single, amodal system mediates the interactions among perceptual representations of different modalities^{5,45,46}. Moreover, observations have suggested that semantic dementia patients experience increasing overregularization of their conceptual knowledge due to the progressive deterioration of this amodal system⁴⁷. Consistent with this, the present study provides robust evidence, through the application of HD-tDCS and TMS, that the integration hubs for gesture and speech—the pMTG and IFG—operate in an incremental manner. This is supported by the progressive inhibition effect observed in these brain areas as the entropy and mutual information of gesture and speech advances.

Moreover, by dividing the potential integration period into eight time windows (TW) relative to the speech identification point (IP) and administering inhibitory double-pulse TMS across each TW, the current study attributed the gradual TMS-selective regional inhibition to distinct information sources. In TW2 of gesture-speech integration, which precedes the speech identification point²³ and represents a pre-lexical stage, the suppression effect observed in the pMTG was correlated with speech entropy. Conversely, during TW6, which follows the speech identification point²³ and represents a post-lexical stage, the IFG interruption effect was influenced by both gesture entropy, speech entropy, and their MI. A dual-stage pMTG-IFG-pMTG neurocircuit loop during gesture-speech integration has been proposed previously²³. As an extension, the present study unveils a staged accumulation of engagement within the neurocircuit linking the transmodal regions of pMTG and IFG, arising from distinct contributors of information.

Furthermore, we disentangled the sub-processes of integration with high-temporal ERPs, when representations of gesture and speech were variously presented. Early P1-N1 and P2 sensory effects linked to perception and attentional processes^{30,48} was comprehended as a reflection of the early audiovisual gesture-speech integration in the sensory-perceptual processing chain⁴⁹. Note that a semantic priming paradigm was adopted here to create a top- down prediction of gesture over speech. The observed positive correlation of the P1 effect with speech entropy and the negative correlation of the N1-P2 effect with gesture entropy suggest that the early interaction of gesture-speech information was modulated by both top- down gesture prediction and bottom-up speech processing. Additionally, the lexico-semantic effect of the N400 and the LPC were differentially mediated by top-down gesture prediction, bottom-up speech encoding and their interaction: the N400 was negatively correlated with both the gesture entropy and MI, but the LPC was negatively correlated only with the speech entropy.

The varying contributions of unisensory gesture-speech information and the convergence of multisensory inputs, as reflected in the correlation between distinct ERP components and TMS time windows (TMS TWs), are consistent with recent models suggesting that multisensory processing involves parallel detection of modality-specific information and hierarchical integration across multiple neural levels^{4,50}. These processes are further characterized by coordination across multiple temporal scales⁵¹. Building on this, the present study offers additional evidence that the multi-level nature of gesture-speech processing is statistically structured, as measured by information matrix of unisensory entropy and multisensory convergence index of MI, the input of either source would activate a distributed representation, resulting in progressively functioning neural responses.

Given that control processes are intrinsically integrated with semantic processing⁵², a distributed semantic representation enables dynamic modulation of access to and manipulation of meaningful information, thereby facilitating flexible control over the diverse possibilities inherent in a concept. Accordingly, an increased number of candidate responses amplifies the control demands necessary to resolve competing semantic representations. This effect was observed in the present study, where the association of the information matrix with the tDCS effect in IFG, the inhibition of pMTG activity in TW2, disruption of IFG activity in TW6, and modulation of four distinct ERP components collectively demonstrated that response quantity modulates neural activity. These results underscore the intricate interplay between the difficulty of semantic representation and the control pressures that shape the resulting neural responses.

The IFG and pMTG, central components of the semantic control network, have been extensively implicated in previous research⁵²⁻⁵⁴. While the role of the IFG in managing both unisensory information and multisensory convergence remains consistent, as evidenced by the confounding difficulty results across Experiments 1 and 2, the current study highlights a more context-dependent function for the pMTG. Specifically, although the pMTG is well- established in the processing of distributed speech information, the multisensory convergence, indexed by MI, did not evoke the same control-related modulation in pMTG activity. These findings suggest that, while the pMTG is critical to semantic processing, its engagement in control processes is likely modulated by the specific nature of the sensory inputs involved.

Considering the close alignment of the ERP components with the TWs of TMS effect, it is reasonable to speculate the ERP components with the cortical involvements (**Figure 6**). Consequently, referencing the recurrent neurocircuit connecting the left IFG and pMTG for semantic unification⁵⁵, we extended the previously proposed two-stage gesture-speech integration circuit²³ into sequential steps. First, bottom-up speech processing mapping acoustic signal to its lexical representation was performed to the pMTG. The larger speech entropy was, the greater effort was made during the matching of the acoustic input with its stored lexical representation, thus leading to a larger involvement of the pMTG at pre-lexical stage (TW2) and a larger P1 effect (**Figure 6**). Second, the gesture representation was activated in the pMTG and

further exerted a top-down modulation over the phonological processing of speech⁵⁶. The higher the certainty of gesture is, a larger modulation of gesture would be made upon speech, as indexed by a smaller gesture entropy with an enhanced N1-P2 amplitude (**Figure 6②**). Third, information was relayed from the pMTG to the IFG for sustained activation, during which a semantic constraint from gesture has been made on the semantic retrieval of speech. Greater TMS inhibitory effect over the IFG at post-lexical stage (TW6) accompanying with a reduced N400 amplitude were found with the increase of gesture entropy, when the representation of gesture was wildly distributed and the constrain over the following speech was weak (**Figure 6③**). Fourth, the activated speech representation was compared with that of the gesture in the IFG. At this stage, the larger, overlapped neural populations activated by gesture and speech as indexed by a larger MI is, a greater TMS disruption effect of the IFG and a reduced N400 amplitude indexing easier integration and less semantic conflict were observed (**Figure 6④**). Last, the activated speech representation would disambiguate and reanalyze the semantic information and further unify into a coherent comprehension in the pMTG^{12,39}. As speech entropy increases, indicating greater uncertainty in the information provided by speech, more cognitive effort is directed towards selecting the targeted semantic representation. This leads to enhanced involvement of the IFG and a corresponding reduction in LPC amplitude (**Figure 6⑤**).

Note that the sequential cortical involvement and ERP components discussed above are derived from a deliberate alignment of speech onset with gesture DP, creating an artificial priming effect with gesture semantically preceding speech. Caution is advised when generalizing these findings to the spontaneous gesture-speech relationships, although gestures naturally precede speech³⁶. Furthermore, MI quantifies overlap in gesture-speech integration, primarily when gestures convey redundant meaning. Consequently, the conclusions drawn in this study are constrained to contexts in which gestures serve to reinforce the meaning of the speech. Future research should aim to explore the neural responses in cases where gestures convey supplementary, rather than redundant, semantic information.

Limitations exist. ERP components and cortical engagements were linked through intermediary variables of entropy and MI. Dissociations were observed between ERP components and cortical engagement. Importantly, there is no direct evidence of the brain structures underpinning the corresponding ERPs, necessitating clarification in future studies. Additionally, not all influenced TWs exhibited significant associations with entropy and MI. While HD-tDCS and TMS may impact functionally and anatomically connected brain regions^{57,58}, whether the absence of influence in certain TWs can be attributed to compensation by other connected brain areas, such as angular gyrus⁵⁹ or anterior temporal lobe⁶⁰, warrants further investigation. Therefore, caution is needed when interpreting the causal relationship between inhibition effects of brain stimulation and information-theoretic metrics (entropy and MI). Finally, the current study incorporated a restricted set of entropy and MI measures. The generalizability of the findings should be assessed in future studies using a more extensive range of matrices.

In summary, utilizing information-theoretic complexity metrics such as entropy and mutual information (MI), our study demonstrates that multisensory semantic processing, involving gesture and speech, gives rise to dynamically evolving representations through the interplay between gesture-primed prediction and speech presentation. This process correlates with the progressive engagement of the pMTG-IFG-pMTG circuit and various ERP components. These findings significantly advancing our understanding of the neural mechanisms underlying multisensory semantic integration.

Appendix Table 2. Examples of ‘an4 (press)’ for the calculation of speech entropy, gesture entropy and mutual information (MI)

Appendix Table 4. Raw RT of semantic congruent (Sc) and semantic incongruent (Si) in Experiment 1 and Experiment 2.

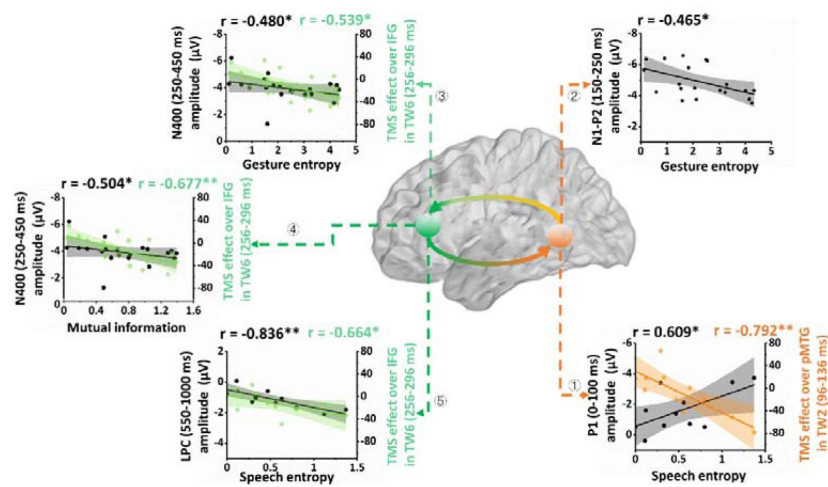


Figure 6.

Progressive processing stages of gesture-speech information within the pMTG-IFG loop.

Correlations between the TMS disruption effect of pMTG and IFG with three information models are represented by the orange line and the green lines, respectively. Black lines denote the strongest correlations of ROI averaged ERP components with three information models. * $p < 0.05$, ** $p < 0.01$ after FDR correction.

Appendix Table 1.

Gesture description and paring with incongruent and congruent speech.

Gesture	Description	Congruent speech	Incongruent speech
an4 (press)	press button	an4 (press)	yun4 (iron)
bai1 (break)	break chopsticks	bai1 (break)	yao2 (shake)
ca1 (wipe)	wipe desk	ca1 (wipe)	reng1 (throw)
ding1 (hammer)	hammer nail	ding1 (hammer)	tui1 (push)
feng2 (sew)	sew cloth	feng2 (sew)	ti2 (lift)
ji3 (squeeze)	squeeze sponge	ji3 (squeeze)	si1 (tear)
jian3 (cut)	cut paper	jian3 (cut)	sao1 (sweep)
jiao3 (stir)	stir flour	jiao3 (stir)	shan1 (slap)
ju4 (saw)	saw wood	ju4 (saw)	ning3 (twist)
ning3 (twist)	twist towel	ning3 (twist)	ju4 (saw)
pen1 (spray)	spray water	pen1 (spray)	qie1 (slice)
qie1 (slice)	slice fruit	qie1 (slice)	pen1 (spray)
reng1 (throw)	throw ball	reng1 (throw)	ca1 (wipe)
shan1 (slap)	slap face	shan1 (slap)	jiao3 (stir)
sao1 (sweep)	sweep floor	sao1 (sweep)	jian3 (cut)
si1 (tear)	tear paper	si1 (tear)	ji3 (squeeze)
ti2 (lift)	lift basket	ti2 (lift)	feng2 (sew)
tui1 (push)	push door	tui1 (push)	ding1 (hammer)
yao2 (shake)	shake bag	yao2 (shake)	bai1 (break)
yun4 (iron)	iron cloth	yun4 (iron)	an4 (press)

Table 2A

Calculation of speech entropy for 'an4.wav (press)'

Answer	Number	p(yi)
a	1	0.016666667
an	33	0.55
e	1	0.016666667
en	23	0.383333333
eng	2	0.033333333
Equation: $\sum_{i=1}^n p(y_i) \log[p(y_i)]$		

Table 2B

Calculation of gesture entropy for 'an4.avi (press)'

Answer	Number	p(xi)
dian	6	0.1
bp	1	0.016666667
chuo	2	0.033333333
dain	1	0.016666667
an	33	0.55
diao	1	0.016666667
en	1	0.016666667
hua	1	0.016666667
shu	3	0.05
zhi	11	0.183333333
Equation: $-\sum_{i=1}^n p(x_i) \log(p(x_i))$		

Table 2C

Calculation of MI for 'an4.avi (press) + an4.wav (press)

Answer	Number	p(xi+yi)
a	1	0.008333333
an	66	0.55
bp	1	0.008333333
chuo	2	0.016666667
dain	1	0.008333333
dian	6	0.05
diao	1	0.008333333
e	1	0.008333333
en	24	0.2
eng	2	0.016666667
hua	1	0.008333333
shu	3	0.025
zhi	11	0.091666667
Equation: $-\sum_{i=1}^n p(x_i) \log(p(x_i)) - \sum_{i=1}^n p(h_i) \log(p(h_i)) - \{-\sum_{i=1}^n p(x + y_i) \log(p(x + y_i))\}$		

Appendix Table 3.

Quantitative information for each stimulus.

Stimuli	Gesture entropy	Speech entropy	Joint entropy	Mutual information	Gesture response No.	Speech response No.	Total response No.
an4 (press)	2.13	1.37	2.15	1.35	10	5	13
bai1 (break)	0.91	0.11	0.61	0.41	5	2	6
ca1 (wipe)	2.07	0.56	1.67	0.96	10	3	10
ding1 (hammer)	2.55	0.00	1.28	1.27	11	1	11
feng2 (sew)	3.04	0.00	1.95	1.09	14	1	14
ji3 (squeeze)	2.50	0.00	1.86	0.64	11	1	11
jian3 (cut)	0.12	0.00	0.07	0.05	2	1	2
jiao3 (stir)	1.83	0.63	1.46	1.01	10	3	12
ju4 (saw)	4.34	0.00	2.77	1.57	25	1	25
ning3 (twist)	0.59	0.63	0.53	0.69	5	3	7
pen1 (spray)	3.29	0.80	2.61	1.49	17	4	20
qie1 (slice)	3.23	0.47	2.31	1.38	18	2	19
reng1 (throw)	1.59	0.29	1.09	0.79	9	2	9
sao1 (sweep)	4.01	1.12	4.47	0.66	23	4	18
shan1 (slap)	1.54	0.33	1.10	0.78	10	3	12
si1 (tear)	0.21	0.00	0.12	0.09	2	1	2
ti2 (lift)	1.48	0.00	0.88	0.60	9	1	9
tui1 (push)	1.62	0.00	0.96	0.66	10	1	17
yao2 (shake)	4.26	0.12	2.93	1.46	25	2	26
yun4 (iron)	4.15	0.00	2.78	1.37	25	1	25

Table 4A

RT of Sc and Si in three HD-tDCS stimulation conditions for IFG and pMTG

		Anodal		Cathodal		Sham	
		Sc (ms) (Rt±SE)	Si (ms) (Rt±SE)	Sc (ms) (Rt±SE)	Si(ms) (Rt±SE)	Sc (ms) (Rt±SE)	Si(ms) (Rt±SE)
tDCS over IFG		521.95 ±13.41	537.46 ±15.05	518.41 ±11.95	530.33 ±13.01	513.96 ±14.40	537.46 ±15.53
tDCS over pMTG		531.94 ±11.43	553.61 ±13.43	531.88 ±11.43	545.08 ±11.97	545.08 ±11.97	569.57 ±14.32

		TW1		TW2		TW3		TW6		TW7	
		Sc (ms) (Rt±SE)	Si (ms) (Rt±SE)	Sc (ms) (Rt±SE)	Si(ms) (Rt±SE)	Sc (ms) (Rt±SE)	Si(ms) (Rt±SE)	Sc(ms) (Rt±SE)	Si(ms) (Rt±SE)	Sc (ms) (Rt±SE)	Si(ms) (Rt±SE)
TMS over	Vertex	507.20 ±12.36	527.06 ±13.44	499.09 ±13.17	534.59 ±15.20	497.65 ±13.99	525.93 ±13.31	497.93 ±13.91	534.46 ±15.85	502.78 ±13.45	524.65 ±11.72
TMS over	IFG	485.11 ±13.80	507.56 ±15.05	486.00 ±13.48	511.71 ±16.01	499.03 ±14.26	507.87 ±15.03	503.21 ±15.32	508.58 ±15.99	490.92 ±14.84	507.38 ±15.71
TMS over	pMTG	498.16 ±15.77	504.78 ±15.10	500.52 ±16.35	510.24 ±16.45	498.42 ±15.26	509.74 ±15.89	497.32 ±15.57	514.01 ±15.87	497.54 ±16.82	502.57 ±16.08

Table 4B

RT of Sc and Si in each time window (TW) for IFG, pMTG, and Vertex

Acknowledgements

This research was supported by grants from the STI 2030—Major Projects 2021ZD0201500, the National Natural Science Foundation of China (31822024, 31800964), the Scientific Foundation of Institute of Psychology, Chinese Academy of Sciences (E2CX3625CX), and the Strategic Priority Research Program of Chinese Academy of Sciences (XDB32010300).

Additional information

Author contributions

Conceptualization, W.Y.Z. and Y.D.; Investigation, W.Y.Z. and Z.Y.L.; Formal Analysis, W.Y.Z. and Z.Y.L.; Methodology, W.Y.Z. and Z.Y.L.; Validation, Z.Y.L. and X.L.; Visualization, W.Y.Z. and Z.Y.L. and X.L.; Funding Acquisition, W.Y.Z. and Y.D.; Supervision, Y.D.; Project administration, Y.D.; Writing – Original Draft, W.Y.Z.; Writing – Review & Editing, W.Y.Z., Z.Y.L., X.L., and Y.D.

Funding

The STI 2030—Major Projects (2021ZD0201500)

National Natural Science Foundation of China (31822024)

National Natural Science Foundation of China (31800964)

Additional files

Appendix Table 1-4 [🔗](#)

References

1. Damasio H., Grabowski T.J., Tranel D., Hichwa R.D., Damasio A.R (1996) **A neural basis for lexical retrieval** *Nature* **380**:499–505 <https://doi.org/10.1038/380499a0> | Google Scholar
2. Patterson K., Nestor P.J., Rogers T.T (2007) **Where do you know what you know? The representation of semantic knowledge in the human brain** *Nature Reviews Neuroscience* **8**:976–987 <https://doi.org/10.1038/nrn2277> | Google Scholar
3. Brennan J.R., Stabler E.P., Van Wagenen S.E., Luh W.M., Hale J.T. (2016) **Abstract linguistic structure correlates with temporal activity during naturalistic comprehension** *Brain and Language* **157**:81–94 <https://doi.org/10.1016/j.bandl.2016.04.008> | Google Scholar
4. Benetti S., Ferrari A., Pavani F (2023) **Multimodal processing in face-to-face interactions: A bridging link between psycholinguistics and sensory neuroscience** *Front Hum Neurosci* **17** <https://doi.org/10.3389/fnhum.2023.1108354> | Google Scholar
5. Noppeney U (2021) **Perceptual Inference, Learning, and Attention in a Multisensory World** *Annual Review of Neuroscience* **44**:449–473 <https://doi.org/10.1146/annurev-neuro-100120-085519> | Google Scholar
6. Ma W.J., Jazayeri M (2014) **Neural coding of uncertainty and probability** *Annu Rev Neurosci* **37**:205–220 <https://doi.org/10.1146/annurev-neuro-071013-014017> | Google Scholar
7. Fischer B.J., Pena J.L (2011) **Owl’s behavior and neural representation predicted by Bayesian inference** *Nat Neurosci* **14**:1061–1066 <https://doi.org/10.1038/nn.2872> | Google Scholar
8. Ganguli D., Simoncelli E.P (2014) **Efficient sensory encoding and Bayesian inference with heterogeneous neural populations** *Neural Comput* **26**:2103–2134 https://doi.org/10.1162/NECO_a_00638 | Google Scholar
9. Hostetter A., Mainela-Arnold E (2015) **Gestures occur with spatial and Motoric knowledge: It’s more than just coincidence** *Perspectives on Language Learning and Education* **22**:42–49 <https://doi.org/10.1044/llc22.2.42> | Google Scholar
10. McNeill D. (2005) **Gesture and thought** University of Chicago Press <https://doi.org/10.7208/chicago/9780226514642.001.0001> | Google Scholar
11. Kendon A (1997) **Gesture** *Annu Rev Anthropol* **26**:109–128 <https://doi.org/10.1146/annurev.anthro.26.1.109> | Google Scholar
12. Hagoort P (2005) **On broca, brain, and binding: a new framework** *Trends in Cognitive Sciences* **9**:416–423 <https://doi.org/10.1016/j.tics.2005.07.004> | Google Scholar
13. Hagoort P., Hald L., Bastiaansen M., Petersson K.M (2004) **Integration of word meaning and world knowledge in language comprehension** *Science* **304**:438–441 <https://doi.org/10.1126/science.1095455> | Google Scholar

14. Ozyurek A., Willems R.M., Kita S., Hagoort P (2007) **On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials** *J Cognitive Neurosci* **19**:605–616 <https://doi.org/10.1162/jocn.2007.19.4.605> | Google Scholar
15. Willems R.M., 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 <https://doi.org/10.1016/j.neuroimage.2009.05.066> | Google Scholar
16. Drijvers L., Jensen O., Spaak E (2021) **Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information** *Human Brain Mapping* **42**:1138–1152 <https://doi.org/10.1002/hbm.25282> | Google Scholar
17. Drijvers L., Ozyurek A (2018) **Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions** *Brain and Language* **177**:7–17 <https://doi.org/10.1016/j.bandl.2018.01.003> | Google Scholar
18. Drijvers L., van der Plas M., Ozyurek A., Jensen O (2019) **Native and non- native listeners show similar yet distinct oscillatory dynamics when using gestures to access speech in noise** *Neuroimage* **194**:55–67 <https://doi.org/10.1016/j.neuroimage.2019.03.032> | Google Scholar
19. Holle H., Gunter T.C (2007) **The role of iconic gestures in speech disambiguation: ERP evidence** *J Cognitive Neurosci* **19**:1175–1192 <https://doi.org/10.1162/jocn.2007.19.7.1175> | Google Scholar
20. 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 [https://doi.org/10.1016/S0749-596x\(02\)00505-3](https://doi.org/10.1016/S0749-596x(02)00505-3) | Google Scholar
21. Bernardis P., Gentilucci M (2006) **Speech and gesture share the same communication system** *Neuropsychologia* **44**:178–190 <https://doi.org/10.1016/j.neuropsychologia.2005.05.007> | Google Scholar
22. Zhao W.Y., Riggs K., Schindler I., Holle H (2018) **Transcranial magnetic stimulation over left inferior frontal and posterior temporal cortex disrupts gesture- speech integration** *Journal of Neuroscience* **38**:1891–1900 <https://doi.org/10.1523/Jneurosci.1748-17.2017> | Google Scholar
23. Zhao W., Li Y., Du Y (2021) **TMS reveals dynamic interaction between inferior frontal gyrus and posterior middle temporal gyrus in gesture-speech semantic integration** *The Journal of Neuroscience* :10356–10364 <https://doi.org/10.1523/jneurosci.1355-21.2021> | Google Scholar
24. Shannon C.E (1948) **A mathematical theory of communication** *Bell Syst Tech J* **27**:379–423 <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x> | Google Scholar
25. Tremblay P., Deschamps I., Baroni M., Hasson U (2016) **Neural sensitivity to syllable frequency and mutual information in speech perception and production** *Neuroimage* **136**:106–121 <https://doi.org/10.1016/j.neuroimage.2016.05.018> | Google Scholar

26. Bikson M., Inoue M., Akiyama H., Deans J.K., Fox J.E., Miyakawa H., Jefferys J.G.R (2004) **Effects of uniform extracellular DC electric fields on excitability in rat hippocampal slices** *J Physiol-London* **557**:175–190 <https://doi.org/10.1113/jphysiol.2003.055772> | Google Scholar
27. Paulus W., Rothwell J.C (2016) **Membrane resistance and shunting inhibition: where biophysics meets state-dependent human neurophysiology** *J Physiol-London* **594**:2719–2728 <https://doi.org/10.1113/jp271452> | Google Scholar
28. Obermeier C., Gunter T.C (2015) **Multisensory integration: The case of a time window of gesture-speech integration** *J Cognitive Neurosci* **27**:292–307 https://doi.org/10.1162/jocn_a_00688 | Google Scholar
29. Federmeier K.D., Mai H., Kutas M (2005) **Both sides get the point: hemispheric sensitivities to sentential constraint** *Memory & Cognition* **33**:871–886 <https://doi.org/10.3758/bf03193082> | Google Scholar
30. Kelly S.D., Kravitz C., Hopkins M (2004) **Neural correlates of bimodal speech and gesture comprehension** *Brain and Language* **89**:253–260 [https://doi.org/10.1016/s0093-934x\(03\)00335-3](https://doi.org/10.1016/s0093-934x(03)00335-3) | Google Scholar
31. Wu Y.C., Coulson S (2005) **Meaningful gestures: Electrophysiological indices of iconic gesture comprehension** *Psychophysiology* **42**:654–667 <https://doi.org/10.1111/j.1469-8986.2005.00356.x> | Google Scholar
32. 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** <https://doi.org/10.1016/j.jml.2020.104191> | Google Scholar
33. Gunter T.C., Weinbrenner J.E.D (2017) **When to take a gesture seriously: On how we use and prioritize communicative cues** *J Cognitive Neurosci* **29**:1355–1367 https://doi.org/10.1162/jocn_a_01125 | Google Scholar
34. Oldfield R.C (1971) **The assessment and analysis of handedness: the Edinburgh inventory** *Neuropsychologia* **9**:97–113 [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4) | Google Scholar
35. Zhao W (2023) **TMS reveals a two-stage priming circuit of gesture-speech integration** *Front Psychol* **14** <https://doi.org/10.3389/fpsyg.2023.1156087> | Google Scholar
36. McNeill D (1992) **Hand and mind : what gestures reveal about thought** University of Chicago Press [Google Scholar](#)
37. Kelly S.D., Creigh P., Bartolotti J (2010) **Integrating speech and iconic gestures in a Stroop-like task: Evidence for automatic processing** *Journal of Cognitive Neuroscience* **22**:683–694 <https://doi.org/10.1162/jocn.2009.21254> | Google Scholar
38. Koessler L., Maillard L., Benhadid A., Vignal J.P., Felblinger J., Vespignani H., Braun M (2009) **Automated cortical projection of EEG sensors: Anatomical correlation via the international 10-10 system** *Neuroimage* **46**:64–72 <https://doi.org/10.1016/j.neuroimage.2009.02.006> | Google Scholar
39. Tesink C.M.J.Y., Petersson K.M., van Berkum J.J.A., van den Brink D, Buitelaar J.K., Hagoort P (2009) **Unification of speaker and meaning in language comprehension: An fMRI study** *J Cognitive Neurosci* **21**:2085–2099 <https://doi.org/10.1162/jocn.2008.21161> | Google Scholar

40. Nuwer M.R., Comi G., Emerson R., Fuglsang-Frederiksen A., Guerit J.M., Hinrichs H., Ikeda A., Luccas F.J., Rappelsberger P (1999) **IFCN standards for digital recording of clinical EEG. The International Federation of Clinical Neurophysiology** *Electroencephalogr Clin Neurophysiol Suppl* **52**:11–14 [https://doi.org/10.1016/S00134694\(97\)00106-5](https://doi.org/10.1016/S00134694(97)00106-5) | Google Scholar
41. Delorme A., Makeig S (2004) **EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis** *J Neurosci Methods* **134**:9–21 <https://doi.org/10.1016/j.jneumeth.2003.10.009> | Google Scholar
42. Habets B., Kita S., Shao Z.S., Ozyurek A., Hagoort P (2011) **The Role of Synchrony and Ambiguity in Speech-Gesture Integration during Comprehension** *J Cognitive Neurosci* **23**:1845–1854 <https://doi.org/10.1162/jocn.2010.21462> | Google Scholar
43. Oostenveld R., Fries P., Maris E., Schoffelen J.-M (2011) **FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data** *Computational Intelligence and Neuroscience* **2011** <https://doi.org/10.1155/2011/156869> | Google Scholar
44. Kutas M., Federmeier K.D (2011) **Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP)** *Annual Review of Psychology* **62**:621–647 <https://doi.org/10.1146/annurev.psych.093008.131123> | Google Scholar
45. Rogers T.T., Ralph M.A.L., Garrard P., Bozeat S., McClelland J.L., Hodges J.R., Patterson K (2004) **Structure and deterioration of semantic memory: A neuropsychological and computational investigation** *Psychological Review* **111**:205–235 <https://doi.org/10.1037/0033-295x.111.1.205> | Google Scholar
46. Ralph M.A.L., Jefferies E., Patterson K., Rogers T.T (2017) **The neural and computational bases of semantic cognition** *Nature Reviews Neuroscience* **18**:42–55 <https://doi.org/10.1038/nrn.2016.150> | Google Scholar
47. Rogers T.T., Hodges J.R., Ralph M.A.L., Patterson K (2003) **Object recognition under semantic impairment: The effects of conceptual regularities on perceptual decisions** *Lang Cognitive Proc* **18**:625–662 <https://doi.org/10.1080/01690960344000053> | Google Scholar
48. Fadiga L., Craighero L., Olivier E (2005) **Human motor cortex excitability during the perception of others' action** *Current Opinion in Neurobiology* **15**:213–218 <https://doi.org/10.1016/j.conb.2005.03.013> | Google Scholar
49. Giard M.H., Peronnet F (1999) **Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study** *J Cognitive Neurosci* **11**:473–490 <https://doi.org/10.1162/089892999563544> | Google Scholar
50. Meijer G.T., Mertens P.E.C., Pennartz C.M.A., Olcese U., Lansink C.S (2019) **The circuit architecture of cortical multisensory processing: Distinct functions jointly operating within a common anatomical network** *Prog Neurobiol* **174**:1–15 <https://doi.org/10.1016/j.pneurobio.2019.01.004> | Google Scholar
51. Senkowski D., Engel A.K (2024) **Multi-timescale neural dynamics for multisensory integration** *Nat Rev Neurosci* **25**:625–642 <https://doi.org/10.1038/s41583-024-00845-7> | Google Scholar

52. Jackson R.L (2021) **The neural correlates of semantic control revisited** *Neuroimage* **224** <https://doi.org/10.1016/j.neuroimage.2020.117444> | Google Scholar
53. Jefferies E (2013) **The neural basis of semantic cognition: converging evidence from neuropsychology, neuroimaging and TMS** *Cortex* **49**:611–625 <https://doi.org/10.1016/j.cortex.2012.10.008> | Google Scholar
54. Noonan K.A., Jefferies E., Visser M., Lambon Ralph M.A (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 https://doi.org/10.1162/jocn_a_00442 | Google Scholar
55. Hagoort P (2013) **MUC (Memory, Unification, Control) and beyond** *Frontiers in Psychology* **4** <https://doi.org/10.3389/fpsyg.2013.00416> | Google Scholar
56. Bizley J.K., Maddox R.K., Lee A.K.C (2016) **Defining auditory-visual objects: Behavioral tests and physiological mechanisms** *Trends in Neurosciences* **39**:74–85 <https://doi.org/10.1016/j.tins.2015.12.007> | Google Scholar
57. 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** <https://doi.org/10.7554/eLife.25964> | Google Scholar
58. Jackson R.L., Hoffman P., Pobric G., Ralph M.A.L (2016) **The semantic network at work and rest: Differential connectivity of anterior temporal lobe subregions** *Journal of Neuroscience* **36**:1490–1501 <https://doi.org/10.1523/JNEUROSCI.2999-15.2016> | Google Scholar
59. Humphreys G.F., Ralph M.A.L., Simons J.S (2021) **A Unifying Account of Angular Gyrus Contributions to Episodic and Semantic Cognition** *Trends in Neurosciences* **44**:452–463 <https://doi.org/10.1016/j.tins.2021.01.006> | Google Scholar
60. Bonner M.F., Price A.R (2013) **Where Is the Anterior Temporal Lobe and What Does It Do?** *Journal of Neuroscience* **33**:4213–4215 <https://doi.org/10.1523/Jneurosci.0041-13.2013> | Google Scholar

Author information

Wanying Zhao

State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China
ORCID iD: [0000-0002-6979-940X](https://orcid.org/0000-0002-6979-940X)

For correspondence: zhaow@psych.ac.cn

Zhouyi Li

State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, School of Psychology, Central China Normal University, Wuhan, China

Xiang Li

State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

Yi Du

State Key Laboratory of Cognitive Science and Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, Chinese Institute for Brain Research, Beijing, China
ORCID iD: [0000-0003-4512-5221](https://orcid.org/0000-0003-4512-5221)

For correspondence: duyi@psych.ac.cn

Editors

Reviewing Editor

Andrea Martin

Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

Senior Editor

Yanchao Bi

Peking University, Beijing, China

Reviewer #1 (Public review):

Summary:

The authors quantified information in gesture and speech, and investigated the neural processing of speech and gestures in pMTG and LIFG, depending on their informational content, in 8 different time-windows, and using three different methods (EEG, HD-tDCS and TMS). They found that there is a time-sensitive and staged progression of neural engagement that is correlated with the informational content of the signal (speech/gesture).

Strengths:

A strength of the paper is that the authors attempted to combine three different methods to investigate speech-gesture processing.

Comments on revisions:

I thank the authors for their careful responses to my comments. However, I remain not convinced by their argumentation regarding the specificity of their spatial targeting and the time-windows that they used.

I do not believe the authors have adequately demonstrated the spatial and temporal specificity required to disentangle the contributions of the IFG and pMTG during the gesture-speech integration process. While the authors have made a sincere effort to address the concerns raised by the reviewers, and have done so with a lot of new analyses, I remain doubtful that the current methodological approach is sufficient to draw conclusions about the causal roles of the IFG and pMTG in gesture-speech integration.

<https://doi.org/10.7554/eLife.99416.3.sa2>

Reviewer #2 (Public review):**Summary**

The study is an innovative and fundamental study that clarified important aspects of brain processes for integration of information from speech and iconic gesture (i.e., gesture that depicts action, movement, and shape), based on tDCS, TMS and EEG experiments. They evaluated their speech and gesture stimuli in information-theoretic ways and calculated how informative speech is (i.e., entropy), how informative gesture is, and how much shared information speech and gesture encode. The tDCS and TMS studies found that the left IFG and pMTG, the two areas that were activated in fMRI studies on speech-gesture integration in the previous literature, are causally implicated in speech-gesture integration. The size of tDC and TMS effects are correlated with entropy of the stimuli or mutual information, which indicates that the effects stems from the modulation of information decoding/integration processes. The EEG study showed that various ERP (event-related potential, e.g., N1-P2, N400, LPC) effects that have been observed in speech-gesture integration experiments in the previous literature are modulated by the entropy of speech/gesture and mutual information. This makes it clear that these effects are related to information decoding processes. The authors propose a model of how speech-gesture integration process unfolds in time, and how IFG and pMTG interact with each other in that process.

Strengths:

The key strength of this study is that the authors used information-theoretic measures of their stimuli (i.e., entropy and mutual information between speech and gesture) in all of their analyses. This made it clear that the neuro-modulation (tDCS, TMS) affected information decoding/integration and ERP effects reflect information decoding/integration. This study used tDCS and TMS methods to demonstrate that left IFG and pMTG are causally involved in speech-gesture integration. The size of tDCS and TMS effects are correlated with information-theoretic measures of the stimuli, which indicate that the effects indeed stem from disruption/facilitation of information decoding/integration process (rather than generic excitation/inhibition). The authors' results also showed correlation between information-theoretic measures of stimuli with various ERP effects. This indicates that these ERP effects reflect the information decoding/integration process.

Weaknesses:

The "mutual information" cannot capture all types of interplay of the meaning of speech and gesture. The mutual information is calculated based on what information can be decoded from speech alone and what information can be decoded from gesture alone. However, when speech and gesture are combined, a novel meaning can emerge, which cannot be decoded from a single modality alone. When example, a person produce a gesture of writing something with a pen, while saying "He paid". The speech-gesture combination can be interpreted as "paying by signing a cheque". It is highly unlikely that this meaning is decoded when people hear speech only or see gestures only. The current study cannot address how such speech-gesture integration occur in the brain, and what ERP effects may reflect such a process. The future studies can classify different types of speech-gesture integration and investigate neural processes that underlie each type. Another important topic for future studies is to investigate how the neural processes of speech-gesture integration change when the relative timing between the speech stimulus and the gesture stimulus changes.

Comments on the previous round of revisions: The authors addressed my concerns well.

<https://doi.org/10.7554/eLife.99416.3.sa1>

Author response:

The following is the authors' response to the previous reviews

Reviewer #1:

Comment:

The authors quantified information in gesture and speech, and investigated the neural processing of speech and gestures in pMTG and LIFG, depending on their informational content, in 8 different time-windows, and using three different methods (EEG, HD-tDCS and TMS). They found that there is a time-sensitive and staged progression of neural engagement that is correlated with the informational content of the signal (speech/gesture).

Strengths:

A strength of the paper is that the authors attempted to combine three different methods to investigate speech-gesture processing.

We sincerely appreciate the reviewer's recognition of our efforts in employing a multi-method approach, which integrates three complementary experimental paradigms, each leveraging distinct neurophysiological techniques to provide converging evidence.

In Experiment 1, we found that the degree of inhibition in the pMTG and LIFG was strongly associated with the overlap in gesture-speech representations, as quantified by mutual information. Experiment 2 revealed the time-sensitive dynamics of the pMTG-LIFG circuit in processing both unisensory (gesture or speech) and multisensory information. Experiment 3, utilizing high-temporal-resolution EEG, independently replicated the temporal dynamics of gesture-speech integration observed in Experiment 2, further validating our findings.

The striking convergence across these methodologically independent approaches significantly bolsters the robustness and generalizability of our conclusions regarding the neural mechanisms underlying multisensory integration.

Comment 1: I thank the authors for their careful responses to my comments. However, I remain not convinced by their argumentation regarding the specificity of their spatial targeting and the time-windows that they used.

The authors write that since they included a sham TMS condition, that the TMS selectively disrupted the IFG-pMTG interaction during specific time windows of the task related to gesture-speech semantic congruency. This to me does not show anything about the specificity of the time-windows itself, nor the selectivity of targeting in the TMS condition.

(1) Selection of brain regions (IFG/pMTG)

We thank the reviewer for their thoughtful consideration. The choice of the left IFG and pMTG as regions of interest (ROIs) was informed by a meta-analysis of fMRI studies on gesture-speech integration, which consistently identified these regions as critical hubs (see Author response table 1 for detailed studies and coordinates).

Author response table 1.

Meta-analysis of previous studies on gesture-speech integration.

Holle et al. 2010, Table1, GSgoodNGsmod, n=16, MNI			
51	-33	3	rMTG/STG
-48	-39	3	LMTG
Green et al. 2009, Table1, German unrelated>German iconic, n=16, MNI			
-60	16	24	LIFG
-60	-60	-4	LMTG
-44	-36	40	l inferior parietal lobule
48	32	12	r inferior frontal gyrus
52	-36	36	r supramarginal gyrus
8	20	56	r/l vsupplementary motor area
44	12	48	r middle frontal gyrus
8	-36	56	r paracental lobule
-20	-12	-16	l hippocampus
Straube et al. 2010, Table3, IC>SNIIC>GIIIGNS, n=17, MNI			
-56	-52	12	LMTG
56	-36	12	RSTG
Willems et al. 2006, Table3, Gmismatch>Gcorrect(global), n=16, Talairach			
-46.46	28.67	26.54	LIFG
-32.32	-48.96	31.15	left intraparietal sulcus(anterior)
-19.2	-66.52	31.32	left intraparietal sulcus(posterior)
Willems et al. 2006, Table3, Lmismatch>Lcorrect(global), n=16, Talairach			
-43	12	24	l inferior frontal sulcus
-33	-65	35	l intraparietal sulcus(posterior)
-52	-50	4	l superior temporal sulcus
-52	-56	6	LpSTS
58	-36	18	RpSTS
-46	10	24	LIFG
-46	22	24	LIFG
14	-92	-2	r inferior occipital sulcus
4	-92	-2	r inferior occipital sulcus
-24	-76	-10	l inferior occipital sulcus
-48	-54	8	l pSTS
62	-36	18	r pSTS
58	-52	10	r pSTS
-40	10	26	l IFG
Holle et al. 2008, Table2, Dominant>Grooming, n=17, MNI			
-8	45	36	left medial middle frontal gyrus
49	3	36	right precentral sulcus
-47	3	33	left precentral sulcus
58	-36	30	right inferior parietal lobule
-59	-36	33	left inferior parietal lobule
34	-39	42	right fusiform gyrus
-50	-54	15	left posterior STS
-38	-75	24	lateral middle occipital gyrus
37	-48	-6	right fusiform gyrus
-8	-96	9	left medial middle occipital gyrus
Holle et al. 2008, Table2, Subordinate>Grooming, n=17, MNI			
-47	6	27	left precentral sulcus
55	-24	39	r inferior parietal lobule
-56	-36	33	left inferior parietal lobule
43	0	27	r precentral sulcus
-41	-51	-9	l fusiform gyrus
-53	-72	12	occipito-temporal junction
40	-48	-9	right fusiform gyrus
-47	-57	12	posterior STS
-54	-59	1	left MTG
-14	-95	-12	l Lingual gyrus
-53	20	18	LIFG

Based on the meta-analysis of previous studies, we selected the IFG and pMTG as ROIs for gesture-speech integration. The rationale for selecting these brain regions is outlined in the introduction in Lines 63-66: "Empirical studies have investigated the semantic integration between gesture and speech by manipulating their semantic relationship[15-18] and revealed a mutual interaction between them19-21 as reflected by the N400 latency and amplitude14 as

well as common neural underpinnings in the left inferior frontal gyrus (IFG) and posterior middle temporal gyrus (pMTG)[15,22,23].”

And further described in Lines 77-78: “Experiment 1 employed high-definition transcranial direct current stimulation (HD-tDCS) to administer Anodal, Cathodal and Sham stimulation to either the IFG or the pMTG”. And Lines 85-88: ‘Given the differential involvement of the IFG and pMTG in gesture-speech integration, shaped by top-down gesture predictions and bottom-up speech processing [23], Experiment 2 was designed to assess whether the activity of these regions was associated with relevant informational matrices”.

In the Methods section, we clarified the selection of coordinates in Lines 194-200: “Building on a meta-analysis of prior fMRI studies examining gesture-speech integration[22], we targeted Montreal Neurological Institute (MNI) coordinates for the left IFG at (-62, 16, 22) and the pMTG at (-50, -56, 10). In the stimulation protocol for HD-tDCS, the IFG was targeted using electrode F7 as the optimal cortical projection site[36], with four return electrodes placed at AF7, FC5, F9, and FT9. For the pMTG, TP7 was selected as the cortical projection site[36], with return electrodes positioned at C5, P5, T9, and P9.”

The selection of IFG or pMTG as integration hubs for gesture and speech has also been validated in our previous studies. Specifically, Zhao et al. (2018, J. Neurosci) applied TMS to both areas. Results demonstrated that disrupting neural activity in the IFG or pMTG via TMS selectively impaired the semantic congruency effect (reaction time costs due to semantic incongruence), while leaving the gender congruency effect unaffected.

These findings identified the IFG and pMTG as crucial hubs for gesture-speech integration, guiding the selection of brain regions for our subsequent studies.

(2) Selection of time windows

The five key time windows (TWs) analyzed in this study were derived from our previous TMS work (Zhao et al., 2021, J. Neurosci), where we segmented the gesture-speech integration period (0–320 ms post-speech onset) into eight 40-ms windows. This interval aligns with established literature on gesture-speech integration, particularly the 200–300 ms window noted by the reviewer. As detailed in Lines (776-779): “Procedure of Experiment 2. Eight time windows (TWs, duration = 40 ms) were segmented in relative to the speech IP. Among the eight TWs, five (TW1, TW2, TW3, TW6, and TW7) were chosen based on the significant results in our prior study[23]. Double-pulse TMS was delivered over each of the TW of either the pMTG or the IFG”.

In our prior work (Zhao et al., 2021, J. Neurosci), we employed a carefully controlled experimental design incorporating two key factors: (1) gesture-speech semantic congruency (serving as our primary measure of integration) and (2) gesture-speech gender congruency (implemented as a matched control factor). Using a time-locked, double-pulse TMS protocol, we systematically targeted each of the eight predefined time windows (TWs) within the left IFG, left pMTG, or vertex (serving as a sham control condition). Our results demonstrated that 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. Crucially, no significant effects were observed during either sham stimulation or the controlled gender congruency factor (Figure 3 from Zhao et al., 2021, J. Neurosci).

This triple dissociation - showing effects only for semantic integration, only in active stimulation, and only at specific time points - provides compelling causal evidence that IFG-pMTG connectivity plays a temporally precise role in gesture-speech integration.

Noted that this work has undergone rigorous peer review by two independent experts who both endorsed our methodological approach. Their original evaluations, provided below:

Reviewer 1: “significance: Using chronometric TMS-stimulation the data of this experiment suggests a feedforward information flow from left pMTG to left IFG followed by an information flow from left IFG back to the left pMTG. The study is the first to provide causal evidence for the temporal dynamics of the left pMTG and left IFG found during gesture-speech integration.”

Reviewer 2: “Beyond the new results the manuscript provides regarding the chronometrical interaction of the left inferior frontal gyrus and middle temporal gyrus in gesture-speech interaction, the study more basically shows the possibility of unfolding temporal stages of cognitive processing within domain-specific cortical networks using short-time interval double-pulse TMS. Although this method also has its limitations, a careful study planning as shown here and an appropriate discussion of the results can provide unique insights into cognitive processing.”

References:

- Willems, R.M., Ozyurek, A., and Hagoort, P. (2009). Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage* 47, 1992-2004. 10.1016/j.neuroimage.2009.05.066.
- Drijvers, L., Jensen, O., and Spaak, E. (2021). Rapid invisible frequency tagging reveals nonlinear integration of auditory and visual information. *Human Brain Mapping* 42, 1138-1152. 10.1002/hbm.25282.
- Drijvers, L., and Ozyurek, A. (2018). Native language status of the listener modulates the neural integration of speech and iconic gestures in clear and adverse listening conditions. *Brain and Language* 177, 7-17. 10.1016/j.bandl.2018.01.003.
- Drijvers, L., van der Plas, M., Ozyurek, A., and Jensen, O. (2019). Native and non-native listeners show similar yet distinct oscillatory dynamics when using gestures to access speech in noise. *Neuroimage* 194, 55-67. 10.1016/j.neuroimage.2019.03.032.
- Holle, H., and Gunter, T.C. (2007). The role of iconic gestures in speech disambiguation: ERP evidence. *J Cognitive Neurosci* 19, 1175-1192. 10.1162/jocn.2007.19.7.1175.
- Kita, S., and 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. 10.1016/S0749-596x(02)00505-3.
- Bernardis, P., and Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia* 44, 178-190. 10.1016/j.neuropsychologia.2005.05.007.
- Zhao, W.Y., Riggs, K., Schindler, I., and Holle, H. (2018). Transcranial magnetic stimulation over left inferior frontal and posterior temporal cortex disrupts gesture-speech integration. *Journal of Neuroscience* 38, 1891-1900. 10.1523/Jneurosci.1748-17.2017.
- Zhao, W., Li, Y., and Du, Y. (2021). TMS reveals dynamic interaction between inferior frontal gyrus and posterior middle temporal gyrus in gesture-speech semantic integration. *The Journal of Neuroscience*, 10356-10364. 10.1523/jneurosci.1355-21.2021.
- Hartwigsen, G., Bzdok, D., Klein, M., Wawrzyniak, M., Stockert, A., Wrede, K., Classen, J., and Saur, D. (2017). Rapid short-term reorganization in the language network. *Elife* 6. 10.7554/eLife.25964.
- Jackson, R.L., Hoffman, P., Pobric, G., and Ralph, M.A.L. (2016). The semantic network at work and rest: Differential connectivity of anterior temporal lobe subregions. *Journal of Neuroscience* 36, 1490-1501. 10.1523/JNEUROSCI.2999-15.2016.

Humphreys, G. F., Lambon Ralph, M. A., & Simons, J. S. (2021). A Unifying Account of Angular Gyrus Contributions to Episodic and Semantic Cognition. *Trends in neurosciences*, 44(6), 452–463. <https://doi.org/10.1016/j.tins.2021.01.006>

Bonner, M. F., & Price, A. R. (2013). Where is the anterior temporal lobe and what does it do?. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 33(10), 4213–4215. <https://doi.org/10.1523/JNEUROSCI.0041-13.2013>

Comment 2: It could still equally well be the case that other regions or networks relevant for gesture-speech integration are targeted, and it can still be the case that these timewindows are not specific, and effects bleed into other time periods. There seems to be no experimental evidence here that this is not the case.

The selection of IFG and pMTG as regions of interest was rigorously justified through multiple lines of evidence. First, a comprehensive meta-analysis of fMRI studies on gesture-speech integration consistently identified these regions as central nodes (see response to comment 1). Second, our own previous work (Zhao et al., 2018, JN; 2021, JN) provided direct empirical validation of their involvement. Third, by employing the same experimental paradigm, we minimized the likelihood of engaging alternative networks. Fourth, even if other regions connected to IFG or pMTG might be affected by TMS, the distinct engagement of specific time windows of IFG and pMTG minimizes the likelihood of consistent influence from other regions.

Regarding temporal specificity, our 2021 study (Zhao et al., 2021, JN, see details in response to comment 1) systematically examined the entire 0-320ms integration window and found that only select time windows showed significant effects for gesture-speech semantic congruency, while remaining unaffected during gender congruency processing. This double dissociation (significant effects for semantic integration but not gender processing in specific windows) rules out broad temporal spillover.

Comment 3: To be more specific, the authors write that double-pulse TMS has been widely used in previous studies (as found in their table). However, the studies cited in the table do not necessarily demonstrate the level of spatial and temporal specificity required to disentangle the contributions of tightly-coupled brain regions like the IFG and pMTG during the speech-gesture integration process. pMTG and IFG are located in very close proximity, and are known to be functionally and structurally interconnected, something that is not necessarily the case for the relatively large and/or anatomically distinct areas that the authors mention in their table.

Our methodological approach is strongly supported by an established body of research employing double-pulse TMS (dpTMS) to investigate neural dynamics across both primary motor and higher-order cognitive regions. As documented in Author response table 1, multiple studies have successfully applied this technique to: (1) primary motor areas (tongue and lip representations in M1), and (2) semantic processing regions (including pMTG, PFC, and ATL). Particularly relevant precedents include:

(1) Teige et al. (2018, *Cortex*): Demonstrated precise spatial and temporal specificity by applying 40ms-interval dpTMS to ATL, pMTG, and mid-MTG across multiple time windows (0-40ms, 125-165ms, 250-290ms, 450-490ms), revealing distinct functional contributions from ATL versus pMTG.

(2) Vernet et al. (2015, *Cortex*): Successfully dissociated functional contributions of right IPS and DLPFC using 40ms-interval dpTMS, despite their anatomical proximity and functional connectivity.

These studies confirm double-pulse TMS can discriminate interconnected nodes at short timescales. Our 2021 study further validated this for IFG-pMTG.

Author response table 2.

Double-pulse TMS studies on brain regions over 3-60 ms time interval

Study	TMS protocol	Results
1. Teige et al., 2018 Cortex	Pair of pulses 40 msec apart was applied over ATL, pMTG and mid-MTG (sham) at various timings (0 & 40 ms; 125 & 165 ms; 250 & 290 ms; 450 & 490 ms).	Stimulation of ATL around 150 ms disrupted the efficient retrieval of strong associations, while stimulation of pMTG at the onset of the second word disrupted the retrieval of weak associations.
2. Amemiya et al., 2017 Scientific Reports	Double-pulse TMS on visual motion area V5/hMT+, primary somatosensory cortex (SI) and posterior parietal cortex (PPC; Brodmann areas 7/40) was delivered 20 ms after the onset of tactile motion across the skin with an interpulse interval of 60 ms.	TMS over both SI and V5/hMT+, but not the PPC site, significantly reduced tactile direction discrimination.
3. Muessgens et al., 2016, Journal of Neurophysiology	Double-pulse TMS was applied over left preSMA or left M1 areas at an interpulse interval of 40 ms.	TMS over the left preSMA significantly slowed responses when stimulation was given between chunks, but only if a hand switch toward the contralateral (right) hand was required. PreSMA stimulation within chunks did not interfere with responses. TMS over the left M1 area delayed responses with the contralateral hand, both within and between chunks.
4. Vernet et al., 2015 Cortex	Paired-pulse TMS was applied over right IPS and dorsolateral prefrontal cortex (DLPFC) 70 and 60 ms before image presentation.	Paired-pulse TMS applied over right IPS and DLPFC 70 and 60 ms before image presentation canceled both the perceptual and neurophysiological effects.
5. Pitcher, 2014 Journal of Neuroscience	Double-pulse TMS (dTMS) was delivered over the right posterior superior temporal sulcus (rpSTS) or the right occipital face area (rOFA) at one of five different latencies after the probe stimulus onset: 20–60, 60–100, 100–140, 130–170, and 170–210 ms.	Task performance was selectively impaired when dTMS was delivered over the rpSTS at 60–100 ms and 100–140 ms. In contrast, dTMS delivered over the rOFA impaired task performance only at 60–100 ms.
6. Bardi et al., 2012 Journal of Cognitive Neuroscience	Double-pulse TMS was applied to the Frontal Eye Field (FEF) of either left or right hemisphere at different time windows (0–40, 40–80, 80–120, 120–180 ms) after the onset of the visual stimulus.	A suppression of the Simon effect was observed after stimulation of the FEF for stimuli appearing in the contralateral hemifield when TMS was applied to the left hemisphere at 0–40 and 40–80 ms after stimulus onset. A reduction of the correspondence effect was observed after right FEF TMS for stimuli presented in the left visual hemifield when stimulation was delivered in the 80–120 ms range after stimulus onset.
7. D'Ausilio et al., 2012 Cortex	Double-pulse TMS on Tongue and Lip M1 localization (lips and tongue motor areas) were given 100 ms and 50 ms before stimulus onset with a 50 ms interval.	A double dissociation between labial and dental syllables discrimination was found when TMS was applied either on LipsM1 or TongueM1
8. Pitcher et al., 2010 Journal of Vision	Double-pulse TMS (separated by 10ms) was delivered over each subject's functionally localised occipital face area (OFA) and extrastriate body area (EBA).	For both OFA and EBA, TMS disrupted task performance at two distinct latencies, 40–50 ms after stimulus onset and 100–110ms after stimulus onset.
9. Gagnon et al., 2010 Brain Research	Paired-pulse TMS was applied for inducing transient and safe interferences over the left or right DLPFCs during encoding or retrieval (3-ms inter-pulse interval).	A reduced discrimination rate following TMS over the left DLPFC was observed during encoding, as compared to a sham condition. Both hit and discrimination rates were also reduced after TMS over the right DLPFC, compared to TMS over left DLPFC, during retrieval.
10. Kaila et al., 2008 NeuroReport	Double-pulse TMS was delivered over frontal eye field (FEF) and posterior parietal cortex (PPC) at different times: 0/40, 40/80, 80/120, 120/160, or 160/200 ms after onset of the search array.	Disruption of performance was earlier (0/40 ms) with FEF stimulation than with PPC stimulation (120/160 ms), revealing a clear and substantial temporal dissociation of the involvement of these two areas in conjunction visual search.
11. Pitcher et al., 2008 Journal of Neuroscience	Double-pulse TMS over right occipital face area (rOFA) and right somatosensory cortex (rSC) with 40 ms between-pulses interval was delivered at seven different times from stimulus onset: 20–60, 60–100, 100–140, 130–170, 170–210, 210–250, and 250–290 ms.	Accuracy dropped when TMS pulses were delivered at 60–100 ms at rOFA and at 100–140 and 130–170 ms at rSC.

References:

Teige, C., Mollo, G., Millman, R., Savill, N., Smallwood, J., Cornelissen, P. L., & Jefferies, E. (2018). Dynamic semantic cognition: Characterising coherent and controlled conceptual retrieval through time using magnetoencephalography and chronometric transcranial magnetic stimulation. *Cortex*, 103, 329-349.

Vernet, M., Brem, A. K., Farzan, F., & Pascual-Leone, A. (2015). Synchronous and opposite roles of the parietal and prefrontal cortices in bistable perception: a double-coil TMS-EEG study. *Cortex*, 64, 78-88.

Comment 4: But also more in general: The mere fact that these methods have been used in other contexts does not necessarily mean they are appropriate or sufficient for investigating the current research question. Likewise, the cognitive processes involved in

these studies are quite different from the complex, multimodal integration of gesture and speech. The authors have not provided a strong theoretical justification for why the temporal dynamics observed in these previous studies should generalize to the specific mechanisms of gesture-speech integration..

The neurophysiological mechanisms underlying double-pulse TMS (dpTMS) are well-characterized. While it is established that single-pulse TMS can produce brief artifacts (typically within 0–10 ms) due to transient cortical depolarization (Romero et al., 2019, NC), the dynamics of double-pulse TMS (dpTMS) involve more intricate inhibitory interactions. Specifically, the first pulse increases membrane conductance via GABAergic shunting inhibition, effectively lowering membrane resistance and attenuating the excitatory impact of the second pulse. This results in a measurable reduction in cortical excitability at the paired-pulse interval, as evidenced by suppressed motor evoked potentials (MEPs) (Paulus & Rothwell, 2016, J Physiol). Importantly, this neurophysiological mechanism is independent of cognitive domain and has been robustly demonstrated across multiple functional paradigms.

In our study, we did not rely on previously reported timing parameters but instead employed a dpTMS protocol using a 40-ms inter-pulse interval. Based on the inhibitory dynamics of this protocol, we designed a sliding temporal window sufficiently broad to encompass the integration period of interest. This approach enabled us to capture and localize the critical temporal window associated with ongoing integrative processing in the targeted brain region.

We acknowledge that the previous phrasing may have been ambiguous, a clearer and more detailed description of the dpTMS protocol has now been provided in Lines 88-92: “To this end, we employed chronometric double-pulse transcranial magnetic stimulation, which is known to transiently reduce cortical excitability at the inter-pulse interval[27]. Within a temporal period broad enough to capture the full duration of gesture–speech integration[28], we targeted specific timepoints previously implicated in integrative processing within IFG and pMTG [23].”

References:

Romero, M.C., Davare, M., Armendariz, M. et al. Neural effects of transcranial magnetic stimulation at the single-cell level. *Nat Commun* 10, 2642 (2019). <https://doi.org/10.1038/s41467-019-10638-7>

Paulus W, Rothwell JC. Membrane resistance and shunting inhibition: where biophysics meets state-dependent human neurophysiology. *J Physiol*. 2016 May 15;594(10):2719-28. doi: 10.1113/JP271452. PMID: 26940751; PMCID: PMC4865581.

Obermeier, C., & Gunter, T. C. (2015). Multisensory Integration: The Case of a Time Window of Gesture-Speech Integration. *Journal of Cognitive Neuroscience*, 27(2), 292-307. https://doi.org/10.1162/jocn_a_00688

Comment 5: Moreover, the studies cited in the table provided by the authors have used a wide range of interpulse intervals, from 20 ms to 100 ms, suggesting that the temporal precision required to capture the dynamics of gesture-speech integration (which is believed to occur within 200-300 ms; Obermeier & Gunter, 2015) may not even be achievable with their 40 ms time windows.

Double-pulse TMS has been empirically validated across neurocognitive studies as an effective method for establishing causal temporal relationships in cortical networks, with demonstrated sensitivity at timescales spanning 3-60 m. Our selection of a 40-ms interpulse interval represents an optimal compromise between temporal precision and physiological feasibility, as evidenced by its successful application in dissociating functional contributions

of interconnected regions including ATL/pMTG (Teige et al., 2018) and IPS/DLPFC (Vernet et al., 2015). This methodological approach combines established experimental rigor with demonstrated empirical validity for investigating the precisely timed IFG-pMTG dynamics underlying gesture-speech integration, as shown in our current findings and prior work (Zhao et al., 2021).

Our experimental design comprehensively sampled the 0-320 ms post-stimulus period, fully encompassing the critical 200-300 ms window associated with gesture-speech integration, as raised by the reviewer. Notably, our results revealed temporally distinct causal dynamics within this period: the significantly reduced semantic congruency effect emerged at IFG at 200-240ms, followed by feedback projections from IFG to pMTG at 240-280ms. This precisely timed interaction provides direct neurophysiological evidence for the proposed architecture of gesture-speech integration, demonstrating how these interconnected regions sequentially contribute to multisensory semantic integration.

Comment 6: I do appreciate the extra analyses that the authors mention. However, my 5th comment is still unanswered: why not use entropy scores as a continuous measure?

Analysis with MI and entropy as continuous variables were conducted employing Representational Similarity Analysis (RSA) (Popal et al., 2019). This analysis aimed to build a model to predict neural responses based on these feature metrics.

To capture dynamic temporal features indicative of different stages of multisensory integration, we segmented the EEG data into overlapping time windows (40 ms in duration with a 10 ms step size). The 40 ms window was chosen based on the TMS protocol used in Experiment 2, which also employed a 40 ms time window. The 10 ms step size (equivalent to 5 time points) was used to detect subtle shifts in neural responses that might not be captured by larger time windows, allowing for a more granular analysis of the temporal dynamics of neural activity.

Following segmentation, the EEG data were reshaped into a four-dimensional matrix (42 channels $\times$ 20 time points $\times$ 97 time windows $\times$ 20 features). To construct a neural similarity matrix, we averaged the EEG data across time points within each channel and each time window. The resulting matrix was then processed using the `pdist` function to compute pairwise distances between adjacent data points. This allowed us to calculate correlations between the neural matrix and three feature similarity matrices, which were constructed in a similar manner. These three matrices corresponded to (1) gesture entropy, (2) speech entropy, and (3) mutual information (MI). This approach enabled us to quantify how well the neural responses corresponded to the semantic dimensions of gesture and speech stimuli at each time window.

To determine the significance of the correlations between neural activity and feature matrices, we conducted 1000 permutation tests. In this procedure, we randomized the data or feature matrices and recalculated the correlations repeatedly, generating a null distribution against which the observed correlation values were compared. Statistical significance was determined if the observed correlation exceeded the null distribution threshold ($p < 0.05$). This permutation approach helps mitigate the risk of spurious correlations, ensuring that the relationships between the neural data and feature matrices are both robust and meaningful.

Finally, significant correlations were subjected to clustering analysis, which grouped similar neural response patterns across time windows and channels. This clustering allowed us to identify temporal and spatial patterns in the neural data that consistently aligned with the semantic features of gesture and speech stimuli, thus revealing the dynamic integration of these multisensory modalities across time. Results are as follows:

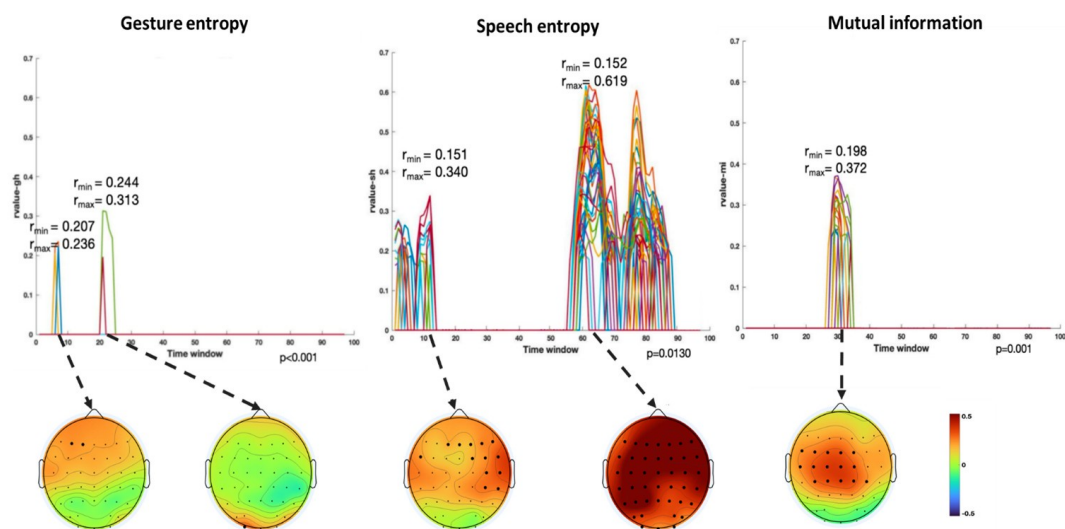
(1) Two significant clusters were identified for gesture entropy (Figure 1 left). The first cluster was observed between 60-110 ms (channels F1 and F3), with correlation coefficients (r) ranging from 0.207 to 0.236 ($p < 0.001$). The second cluster was found between 210-280 ms (channel O1), with r -values ranging from 0.244 to 0.313 ($p < 0.001$).

(2) For speech entropy (Figure 1 middle), significant clusters were detected in both early and late time windows. In the early time windows, the largest significant cluster was found between 10-170 ms (channels F2, F4, F6, FC2, FC4, FC6, C4, C6, CP4, and CP6), with r -values ranging from 0.151 to 0.340 ($p = 0.013$), corresponding to the P1 component (0-100 ms). In the late time windows, the largest significant cluster was observed between 560-920 ms (across the whole brain, all channels), with r -values ranging from 0.152 to 0.619 ($p = 0.013$).

(3) For mutual information (MI) (Figure 1 right), a significant cluster was found between 270-380 ms (channels FC1, FC2, FC3, FC5, C1, C2, C3, C5, CP1, CP2, CP3, CP5, FCz, Cz, and CPz), with r -values ranging from 0.198 to 0.372 ($p = 0.001$).

Author response image 1.

Results of RSA analysis.



These additional findings suggest that even using a different modeling approach, neural responses, as indexed by feature metrics of entropy and mutual information, are temporally aligned with distinct ERP components and ERP clusters, as reported in the current manuscript. This alignment serves to further consolidate the results, reinforcing the conclusion we draw. Considering the length of the manuscript, we did not include these results in the current manuscript.

Reference:

Popal, H., Wang, Y., & Olson, I. R. (2019). A guide to representational similarity analysis for social neuroscience. *Social cognitive and affective neuroscience*, 14(11), 1243-1253.

Comment 7: In light of these concerns, I do not believe the authors have adequately demonstrated the spatial and temporal specificity required to disentangle the contributions of the IFG and pMTG during the gesture-speech integration process. While the authors have made a sincere effort to address the concerns raised by the reviewers, and have done so with a lot of new analyses, I remain doubtful that the current

methodological approach is sufficient to draw conclusions about the causal roles of the IFG and pMTG in gesture-speech integration.

To sum up:

(1) Empirical validation from our prior work (Zhao et al., 2018,2021,JN): The selection of IFG and pMTG as target regions was informed by both: (1) a comprehensive meta-analysis of fMRI studies on gesture-speech integration, and (2) our own prior causal evidence from Zhao et al. (2018, J Neurosci), with detailed stereotactic coordinates provided in the attached Response to Editors and Reviewers letter. The temporal parameters were similarly grounded in empirical data from Zhao et al. (2021, J Neurosci), where we systematically examined eight consecutive 40-ms windows spanning the full integration period (0-320 ms). This study revealed a triple dissociation of effects - occurring exclusively during: (i) semantic integration (but not control tasks), (ii) active stimulation (but not sham), and (iii) specific time windows (but not all time windows)- providing robust causal evidence for the spatiotemporal specificity of IFG-pMTG interactions in gesture-speech processing. Notably, all reviewers recognized the methodological strength of this dpTMS approach in their evaluations (see attached JN assessment for details).

(2) Convergent evidence from Experiment 3: Our study employed a multi-method approach incorporating three complementary experimental paradigms, each utilizing distinct neurophysiological techniques to provide converging evidence. Specifically, Experiment 3 implemented high-temporal-resolution EEG, which independently replicated the time-sensitive dynamics of gesture-speech integration observed in our double-pulse TMS experiments. The remarkable convergence between these methodologically independent approaches -demonstrating consistent temporal staging of IFG-pMTG interactions across both causal (TMS) and correlational (EEG) measures - significantly strengthens the validity and generalizability of our conclusions regarding the neural mechanisms underlying multisensory integration.

(3) Established precedents in double-pulse TMS literature: The double-pulse TMS methodology employed in our study is firmly grounded in established neuroscience research. As documented in our detailed Response to Editors and Reviewers letter (citing 11 representative studies), dpTMS has been extensively validated for investigating causal temporal dynamics in cortical networks, with demonstrated sensitivity at timescales ranging from 3-60 ms. Particularly relevant precedents include: 1. Teige et al. (2018, Cortex) successfully dissociated functional contributions of anatomically proximal regions (ATL vs. pMTG vs.mid-MTG) using 40-ms-interval double-pulse TMS; 2. Vernet et al. (2015, Cortex) effectively distinguished neural processing in interconnected frontoparietal regions (right IPS vs. DLPFC) using 40-ms double-pulse TMS parameters. Both parameters are identical to those employed in our current study.

(4) Neurophysiological Plausibility: The neurophysiological basis for the transient double-pulse TMS effects is well-established through mechanistic studies of TMS-induced cortical inhibition (Romero et al.,2019; Paulus & Rothwell, 2016).

Taking together, we respectfully submit that our methodology provides robust support for our conclusions.

<https://doi.org/10.7554/eLife.99416.3.sa0>