

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

Reviewed Preprint

v1 • December 13, 2024

Not revised

Wanying Zhao , Zhouyi Li, Xiang Li, Yi Du 

CAS Key Laboratory of Behavioral Science, 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 • CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai, China • Chinese Institute for Brain Research, Beijing, China 102206

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

eLife Assessment

This **useful** study 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. It uses brain stimulation and electroencephalography to study speech-gesture integration. While the research question is interesting, the methods are insufficient for studying tightly-coupled brain regions over short timescales, leading to **incomplete** support for the claims due to conceptual and methodological limitations.

<https://doi.org/10.7554/eLife.99416.1.sa4>

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 areas^{1,2}, or the 'semantic hub' located in the anterior temporal lobe², pivotal for integrating, converging, or distilling multimodal inputs. Contemporary perspectives on semantic processing portray it as a sequence of quantitatively functional mental states defined by a specific parser³, unified by statistical regularities among multiple sensory inputs⁴ through hierarchical prediction and multimodal interactions^{5–9}. Hence, proposals suggest that the coherent semantic representation emerges from statistical learning mechanisms within these 'convergence zones' or 'semantic hub'^{10–12}, potentially functioning in a graded manner^{12,13}. However, the exact nature of the graded structure within these integration hubs, along with their temporal dynamics, remains elusive.

Among the many kinds of multimodal extralinguistic information, representational gesture is the one that is related to the semantic content of co-occurring speech^{14,15}. Representational gesture is regarded as 'part of language'¹⁶ or functional equivalents of lexical units that alternate and integrate with speech into a 'single unification space' to convey a coherent meaning^{17–19}. Empirical studies have investigated the semantic integration between representational gesture (gesture in short hereafter) and speech by manipulating their semantic relationship^{20–23} and revealed a mutual interaction between them^{24–26} as reflected by the N400 latency and amplitude¹⁹ as well as common neural underpinnings in the left inferior frontal gyrus (IFG) and posterior middle temporal gyrus (pMTG)^{20,27,28}. Quantifying the amount of information from both sources and their interaction, the present study delved into cortical engagement and temporal dynamics during multisensory gesture-speech integration, with a specific focus on the IFG and pMTG, alongside various ERP components.

To this end, we developed an analytic approach to directly probe the contribution of gesture and speech during multisensory semantic integration, while adopting the information-theoretic complexity metrics of *entropy* and *mutual information* (MI). Entropy captures the disorder or randomness of information and is used as a measurement of the uncertainty of representation activated when an event occurs²⁹. MI illustrates the mutual constraint that the two variables impose on each other³⁰. Herein, during gesture-speech integration, entropy measures the uncertainty of information of gesture or speech, while MI indexes the degree of integration.

Three experiments were conducted to unravel the intricate neural processes underlying gesture-speech semantic integration. In Experiment 1, High-Definition Transcranial Direct Current Stimulation (HD-tDCS) was utilized to administer Anodal, Cathodal and Sham stimulation to either the IFG or the pMTG. HD-tDCS induces membrane depolarization with anodal stimulation and membrane hyperpolarisation with cathodal stimulation³¹, thereby respectively increasing or decreasing cortical excitability in the targeted brain area. Hence, Experiment 1 aimed to determine whether the facilitation effect (Anodal-tDCS minus Sham-tDCS) and/or the inhibitory effect (Cathodal-tDCS minus Sham-tDCS) on the integration hubs of IFG and/or pMTG were modulated by the degree of gesture-speech integration, indexed with MI. Considering the different roles of IFG and pMTG during integration²⁸, as well as the various ERP components reported in prior investigations, such as the early sensory effect as P1 and N1–P2^{33,34}, the N400 semantic conflict effect^{19,34,35}, and the late positive component (LPC) reconstruction effect^{36,37}. Experiment 2 employed chronometric double-pulse transcranial magnetic stimulation (TMS) to target short time windows along the gesture-speech integration period³². In parallel, Experiment 3 utilized high-temporal event-related potentials to explore whether the various neural engagements were temporally and progressively modulated by distinct information contributors during gesture-speech integration.

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)³⁸ = 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 study²⁸ were used. The stimuli set were recorded from two native Chinese speakers (1 male, 1 female) and validated by replicating the semantic congruency effect with 30 participants. Results showed 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)²⁸.

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 pre-test, the response consistently provided by participants for four to six consecutive instances was considered the comprehensive answer 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, 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). The joint entropy computation for gesture and speech, represented by $H(x_i, y_i)$, involved amalgamating datasets of gesture and speech responses to depict their combined distributions. For specific gesture-speech combinations, equivalence between the joint 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 joint entropy and the sum of gesture and speech entropies. This quantification of gesture-speech overlap was operationalized by subtracting the joint entropy of gesture-speech from the combined entropies of gesture and speech, indexed by Mutual Information (MI) (see Appendix Table 2C). Elevated MI values thus signify substantial overlap,

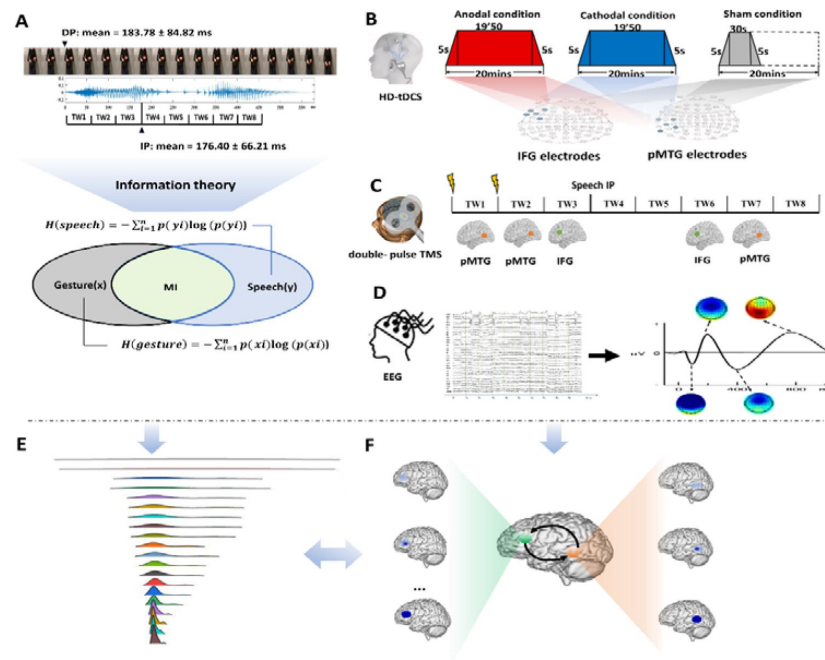


Figure 1.

Experimental design, and stimulus characteristics.

(A) Experimental stimuli. Twenty gestures were paired with 20 relevant speech stimuli. Two gating studies 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.

(B) Procedure 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 study²⁸. 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.

(E-F) Proposed gradations in cortical engagements during gesture-speech information changes. Stepwise variations in the quantity of gesture and speech information during integration, as characterized by information theory metrics (E), are believed to be underpinned by progressive neural engagement within the IFG-pMTG gesture-speech integration circuit (F).

indicative of a robust mutual interaction between gesture and speech. The quantitative information for each stimulus, including gesture entropy, speech entropy, joint entropy, and MI are displayed in **Appendix Table 3** [↗](#).

To accurately assess whether entropy/MI corresponds to stepped neural changes, the current study aggregated neural responses (Non-invasive brain stimulation (NIBS) inhibition effect or ERP amplitude) with identical entropy or MI values prior to conducting correlational analyses.

Experimental procedure

Adopting a semantic priming paradigm of gestures onto speech^{16,32} [↗](#), speech onset was set to be at the DP of each accompanying gesture. An irrelevant factor of gender congruency (e.g., a man making a gesture combined with a female voice) was created^{27,28,39} [↗](#). 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 speech³⁹ [↗](#). In light of previous findings indicating a distinct TMS-disruption effect on the semantic congruency of gesture-speech interactions²⁸ [↗](#), 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 underwent Experiment 1, comprising 480 gesture-speech pairs, across three separate sessions spaced one week apart for each participant. In each session, participants received one of three stimulation types (Anodal, Cathodal, or Sham). Experiment 2 consisted of 800 pairs and was conducted across 15 blocks over three days, with one week between sessions. The order of stimulation site and time window (TW) was counterbalanced using a Latin square design. Experiment 3, comprising 80 gesture-speech pairs, was completed in a single-day session. 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 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

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. For targeting the left IFG at Montreal Neurological Institute (MNI) coordinates (-62, 16, 22), electrode F7 was selected as the optimal cortical projection site⁴⁰ [↗](#), with the four return electrodes placed on AF7, FC5, F9, and FT9. For stimulation of the pMTG at coordinates (-50, -56, 10), TP7 was identified as the cortical projection site⁴⁰ [↗](#), with return electrodes positioned on 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. Our present analysis focused on Pearson correlations between the interruption effects of HD-tDCS (active tDCS minus sham tDCS) on the semantic congruency effect (difference in reaction time between semantic incongruent and semantic congruent pairs) and the variables of

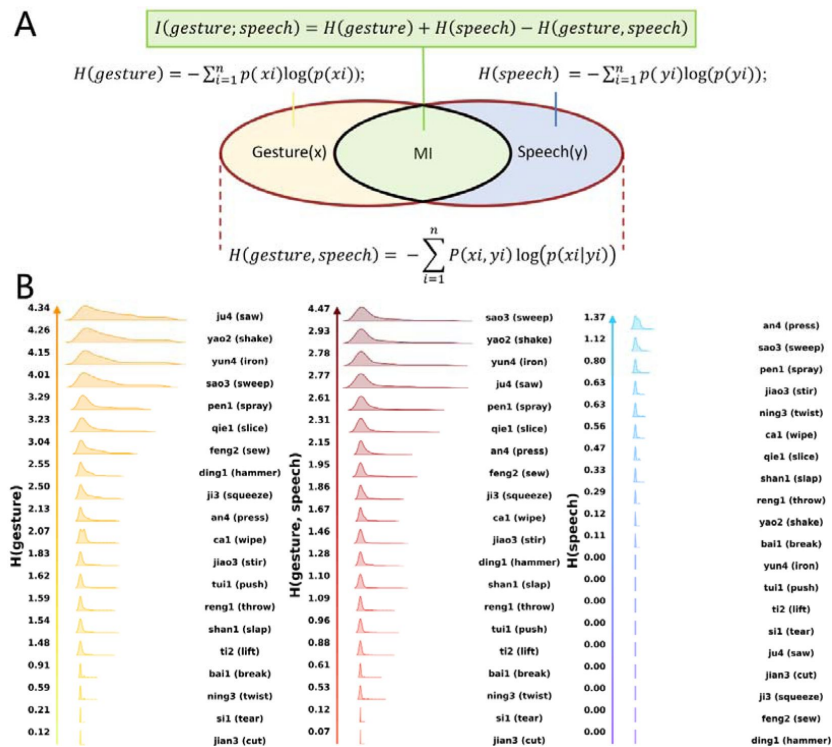


Figure 2.

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.

gesture entropy, speech entropy, or MI. This methodology seeks to determine whether the neural activity within the left IFG and pMTG is gradually affected by varying levels of gesture and speech information during integration, as quantified by entropy and MI.

Experiment 2: TMS protocol and data analysis

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) over either the left IFG in TW3 (-40~0 ms in relative to speech identification point (IP)) and TW6 (80~120 ms,) or the left pMTG in TW1 (-120 ~ -80 ms), TW2 (-80 ~ -40 ms) and TW7 (120~160 ms). Among the TWs that covering the period of gesture-speech integration, those that showed a TW-selective disruption of gesture-speech integration were selected²⁸ (Figure 1C).

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. 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

EEG were recorded from 48 Ag/AgCl electrodes mounted in a cap according to the 10-20 system⁴¹, 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 data⁴². 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 objectively identify the time windows of activated components, 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 predetermined: the P1 effect observed within the time window of 0-100 ms^{33,34}, the N1-P2 effect observed between 150-250ms^{33,34}, the N400 within the interval of 250-450ms^{19,34,35}, and the LPC spanning from 550-1000ms^{36,37}. 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 CPz⁴³.

Subsequently, cluster-based permutation tests⁴⁴ 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 calculated between each model value and each averaged ERP amplitude in each ROI or cluster, individually. False discovery rate (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).

Similarly, 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). 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

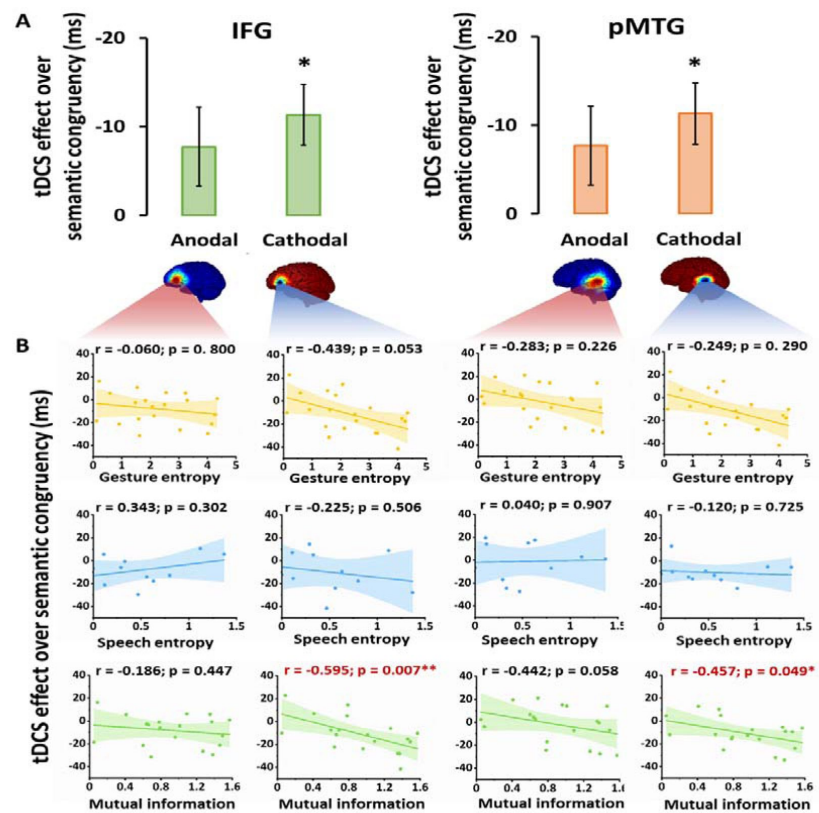


Figure 3.

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.

(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.

($t(23) = -2.335$, $p = 0.029$, 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 more negative TMS effect represents a stronger interruption of the integration effect) 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).

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 5B2 middle). Aligning with prior findings³³, 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 stimuli with higher 50% speech entropy against those with lower 50% speech entropy (Figure 5B2 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 5B2 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 5B3 middle) ($t(22) = 2.754$, $p = 0.004$, 95% confidence interval (CI) = [1.65, 3.86], Figure 5B3 left), with a significant correlation with speech entropy ($r = 0.636$, $p = 0.035$, 95% CI = [0.081, 1.191], Figure 5B3 right).

Additionally, topographical maps comparing the lower 50% and higher 50% gesture entropy revealed a frontal N1-P2 amplitude (150-250 ms, Figure 5A2 middle). In accordance with previous findings on bilateral frontal N1-P2 amplitude³³, 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 5A2 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 5A2 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 5A3 middle) ($t(22) = 2.638$, $p = 0.015$, 95% CI = [1.79, 3.48], (Figure 5A3 left)), exhibiting a significant correlation with gesture entropy ($r = -0.485$, $p = 0.030$, 95% CI = [-0.91, -0.06], Figure 5A3 right).

Furthermore, in line with prior research⁴⁵, a left-frontal N400 amplitude (250-450 ms) was discerned from topographical maps of both gesture entropy (Figure 5A4 middle) and MI (Figure 5C2 middle). Notably, a larger N400 amplitude in the LA ROI was consistently observed for stimuli with lower 50% values compared to those with higher 50% values, both for gesture entropy ($t(22) = 2.455$, $p = 0.023$, 95% CI = [1.95, 2.96], Figure 5A4 left) and MI ($t(22) = 3.00$, $p = 0.007$, 95% CI = [2.54, 3.46], Figure 5C2 left). Concurrently, a negative correlation was noted between the N400 amplitude and both gesture entropy ($r = -0.480$, $p = 0.032$, 95% CI = [-0.94, -0.03], Figure 5A4 right) and MI ($r = -0.504$, $p = 0.028$, 95% CI = [-0.97, -0.04], Figure 5C2 right) in the LA ROI.

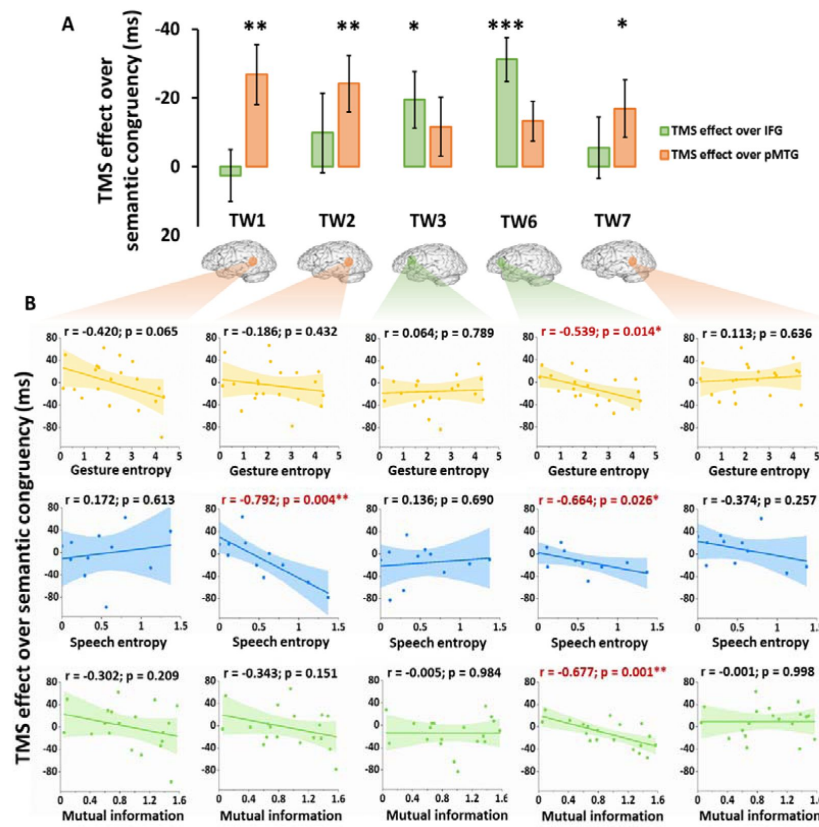


Figure 4.

TMS effect over semantic congruency.

(A) 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.

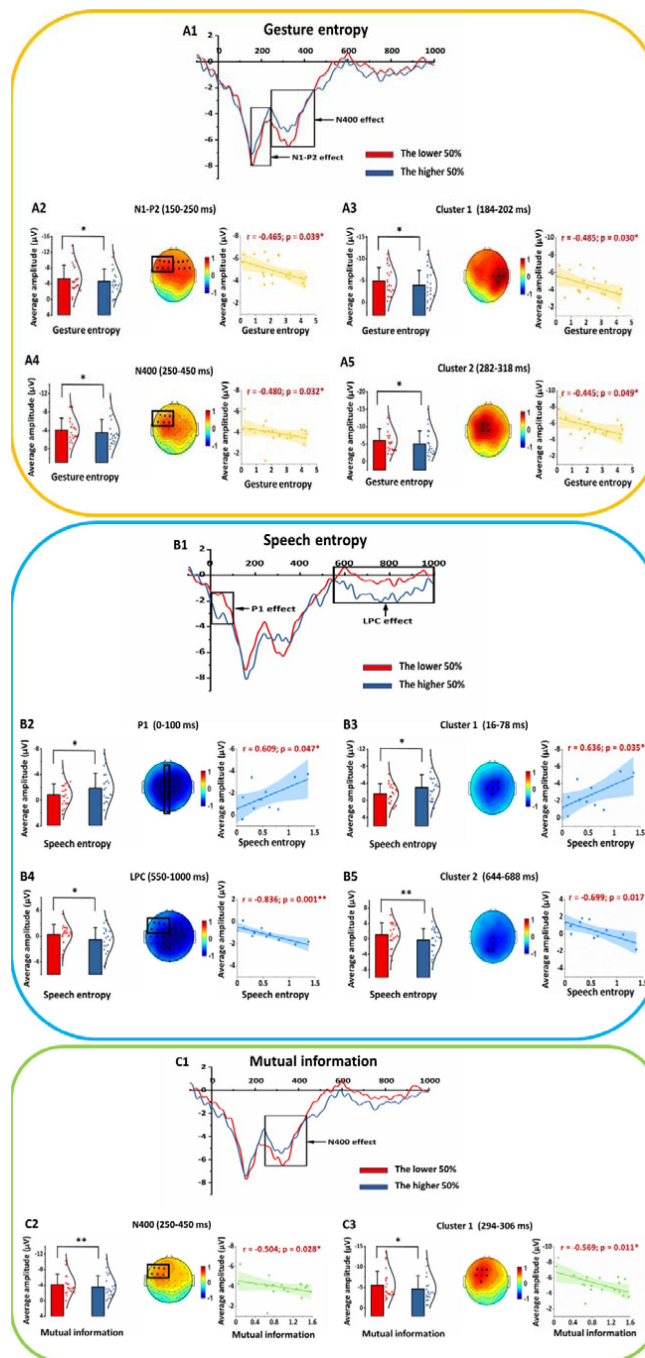


Figure 5.

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 (**Top panels**). Clusters of adjacent time points and electrodes were subsequently identified within each component using a cluster-based permutation test (**Bottom right**). Topographical maps depict amplitude differences between the lower and higher halves of each information model, with significant ROIs or electrode clusters highlighted in black. Solid rectangles delineating the ROIs that exhibited the maximal correlation and paired t-values (**Bottom left**). T-test comparisons with normal distribution lines and correlations with best-fitting regression lines are calculated and illustrated between the average ERP amplitude within the rectangular ROI (**Bottom left**) or the elicited clusters (**Bottom right**) and the three information models individually. * $p < 0.05$, ** $p < 0.01$ after FDR correction.

The identified clusters with the N400 effect for gesture entropy (282 – 318 ms at electrodes FC1, FCz, C1, and Cz, **Figure 5A5 middle**) ($t(22) = 2.828$, $p = 0.010$, 95% CI = [2.02, 3.64], **Figure 5A5 left**) exhibited significant correlation between the N400 amplitude and gesture entropy ($r = -0.445$, $p = 0.049$, 95% CI = [-0.88, -0.01], **Figure 5A5 right**). Similarly, the cluster with the N400 effect for MI (294 – 306 ms at electrodes F1, F3, Fz, FC1, FC3, FCz, and C1, **Figure 5C3 middle**) ($t(22) = 2.461$, $p = 0.023$, 95% CI = [1.62, 3.30], **Figure 5C3 left**) also exhibited significant correlation ($r = -0.569$, $p = 0.011$, 95% CI = [-0.98, -0.16], **Figure 5C5 right**).

Finally, consistent with previous findings³³, an anterior LPC effect (550-1000 ms) was observed in topographical maps comparing stimuli with lower and higher 50% speech entropy (**Figure 5B4 middle**). 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 5B4 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 5B4 right**). Additionally, a cluster with the LPC effect (644 - 688 ms at electrodes Cz, CPz, P1, and Pz, **Figure 5B5 middle**) ($t(22) = 2.754$, $p = 0.012$, 95% CI = [1.50, 4.01], **Figure 5B5 left**) displayed a significant correlation with speech entropy ($r = -0.699$, $p = 0.017$, 95% CI = [-1.24, -0.16], **Figure 5B5 right**).

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, unveiled a progressive inhibition of IFG and pMTG by HD-tDCS as the degree of gesture-speech interaction, indexed by MI, advanced (Experiment 1). Additionally, 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 top-down gesture constrain (gesture entropy) and bottom-up speech. 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^{11,12,46}. Moreover, observations have suggested that semantic dementia patients experience increasing overregularization of their conceptual knowledge due to the progressive deterioration of this amodal system⁴⁷. Consequently, a graded function and structure of the transmodal 'hub' representational system has been proposed^{12,48,49}. In line with this, through the use of NIBS techniques such as HD-tDCS and TMS, the present study provides compelling evidence that the integration hubs of gesture and speech, namely the pMTG and IFG, function in a graded 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 TWs relative to the speech 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 the early pre-lexical TW2 of gesture-speech integration, the suppression effect observed in the pMTG was correlated with speech entropy. Conversely, in the later post-lexical TW6, 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 previously²⁸. 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^{34,50} was comprehended as a reflection of the early audiovisual gesture-speech integration in the sensory-perceptual processing chain⁵¹. 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. Nonetheless, activation of representation is modulated progressively. The input stimuli would activate a dynamically distributed neural landscape, the state of which constructs gradually as measured by entropy and MI and correlates with the electrophysiological signals (N400 and LPC) which indicate the change of lexical representation. Consistent with recent account in multisensory information processing^{4,52}, our findings further confirm that the changed activation pattern can be induced from directions of both top-down and bottom-up gesture-speech processing.

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 unification⁵³, we extended the previously proposed two-stage gesture-speech integration circuit²⁸ into sequential steps. First, bottom-up speech processing mapping acoustic signal to its lexical representation was performed from the STG/S 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** ①). Second, the gesture representation was activated in the pMTG and further exerted a top-down modulation over the phonological processing of speech in the STG/S⁵⁴. The higher certainty of gesture, a larger modulation of gesture would be made upon speech, as indexed by a smaller gesture entropy with an enhanced N1-P2 amplitude (**Figure 6** ②). 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 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 widely distributed and the constrain over the following speech was weak (**Figure 6** ③). 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, a greater TMS disruption effect of the IFG and a reduced N400 amplitude indexing easier integration and less semantic conflict were observed (**Figure 6** ④). Last, the activated speech representation would disambiguate and reanalyze the semantic information that was stored in the IFG and further unify into a coherent comprehension in the pMTG^{17,55}. The more uncertain information being provided by speech, as indicated by an increased speech entropy, a stronger reweighting effect was made over the activated semantic information, resulting in a strengthened involvement of the IFG as well as a reduced LPC amplitude (**Figure 6** ⑤).

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

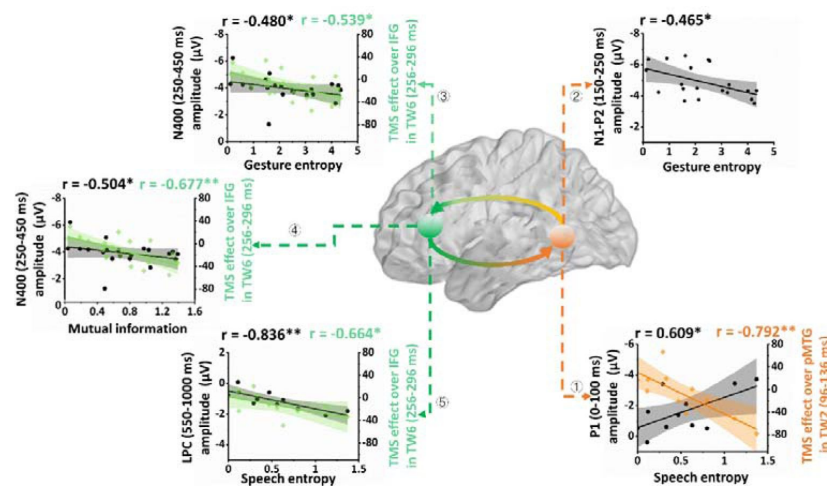


Figure 6.

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.

to the spontaneous gesture-speech relationships, although gestures naturally precede speech⁵⁶ [↗](#).

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 regions⁴³ [↗](#),⁴⁴ [↗](#), the graded functionality of every disturbed period is not guaranteed. Caution is warranted in interpreting the causal relationship between NIBS inhabitation effects 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.

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.

Competing interests

The authors declare no competing interests.

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)

Appendix Table 1.

Gesture description and paring with incongruent and congruent speech.

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:

$$-\sum_{i=1}^n p(x_i) \log(p(x_i)) - \sum_{i=1}^n p(y_i) \log(p(y_i)) - \{-\sum_{i=1}^n p(x_i, y_i) \log(p(x_i, y_i))\}$$

Appendix Table 2.

Examples of 'an4 (press)' for the calculation of speech entropy, gesture entropy and mutual information (MI)

Stimuli	Gesture entropy	Speech entropy	Joint entropy	Mutual information
an4 (press)	2.13	1.37	2.15	1.35
bai1 (break)	0.91	0.11	0.61	0.41
ca1 (wipe)	2.07	0.56	1.67	0.96
ding1 (hammer)	2.55	0.00	1.28	1.27
feng2 (sew)	3.04	0.00	1.95	1.09
ji3 (squeeze)	2.50	0.00	1.86	0.64
jian3 (cut)	0.12	0.00	0.07	0.05
jiao3 (stir)	1.83	0.63	1.46	1.01
ju4 (saw)	4.34	0.00	2.77	1.57
ning3 (twist)	0.59	0.63	0.53	0.69
pen1 (spray)	3.29	0.80	2.61	1.49
qie1 (slice)	3.23	0.47	2.31	1.38
reng1 (throw)	1.59	0.29	1.09	0.79
sao1 (sweep)	4.01	1.12	4.47	0.66
shan1 (slap)	1.54	0.33	1.10	0.78
si1 (tear)	0.21	0.00	0.12	0.09
ti2 (lift)	1.48	0.00	0.88	0.60
tui1 (push)	1.62	0.00	0.96	0.66
yao2 (shake)	4.26	0.12	2.93	1.46
yun4 (iron)	4.15	0.00	2.78	1.37

Appendix Table 3.

Quantitative information for each stimulus.

Appendix Table 4.

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

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

Table 4B: RT of Sc and Si in each time window (TW) for IFG, pMTG, and Vertex

	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

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>
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>
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>
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>
5. Armeni K., Willems R.M., Frank S.L (2017) **Probabilistic language models in cognitive neuroscience: Promises and pitfalls** *Neurosci Biobehav R* **83**:579–588 <https://doi.org/10.1016/j.neubiorev.2017.09.001>
6. Leonard M.K., Bouchard K.E., Tang C., Chang E.F (2015) **Dynamic encoding of speech sequence probability in human temporal cortex** *Journal of Neuroscience* **35**:7203–7214 <https://doi.org/10.1523/JNEUROSCI.4100-14.2015>
7. Goldstein A. *et al.* (2022) **Shared computational principles for language processing in humans and deep language models** *Nature Neuroscience* **25** <https://doi.org/10.1038/s41593-022-01026-4>
8. Heilbron M., Armeni K., Schoffelen J.-M., Hagoort P., de Lange F.P. (2022) **A hierarchy of linguistic predictions during natural language comprehension** *P Natl Acad Sci USA* **119**:e2201968119–e2201968119 <https://doi.org/10.1073/pnas.2201968119>
9. Kuperberg G.R., Jaeger T.F (2016) **What do we mean by prediction in language comprehension?** *Lang Cogn Neurosci* **31**:32–59 <https://doi.org/10.1080/23273798.2015.1102299>
10. Ralph M.A.L (2014) **Neurocognitive insights on conceptual knowledge and its breakdown** *Philos T R Soc B* **369** <https://doi.org/10.1098/rstb.2012.0392>
11. 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>
12. 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>

13. Holler J., Levinson S.C (2019) **Multimodal language processing in human communication** *Trends in Cognitive Sciences* **23**:639–652 <https://doi.org/10.1016/j.tics.2019.05.006>
14. 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/lle22.2.42>
15. McNeill D. (2005) **Gesture and thought** University of Chicago Press <https://doi.org/10.7208/chicago/9780226514642.001.0001>
16. Kendon A (1997) **Gesture** *Annu Rev Anthropol* **26**:109–128 <https://doi.org/10.1146/annurev.anthro.26.1.109>
17. 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>
18. 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>
19. 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** *Cognitive Neurosci* **19**:605–616 <https://doi.org/10.1162/jocn.2007.19.4.605>
20. 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>
21. 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>
22. 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>
23. 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>
24. 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>
25. 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)
26. 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>
27. 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>

28. 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* <https://doi.org/10.1523/jneurosci.1355-21.2021>
29. 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>
30. 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>
31. 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>
32. Pitcher D., Walsh V., Yovel G., Duchaine B (2007) **TMS evidence for the involvement of the right occipital face area in early face processing** *Current Biology* **17**:1568–1573 <https://doi.org/10.1016/j.cub.2007.07.063>
33. 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>
34. 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)
35. 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>
36. 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>
37. 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
38. 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)
39. 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>
40. 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>

41. 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/S0013-4694\(97\)00106-5](https://doi.org/10.1016/S0013-4694(97)00106-5)
42. 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>
43. 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>
44. 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>
45. 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>
46. 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>
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>
48. Krieger-Redwood K., Wang X., Souter N., Gonzalez Alam T.R.d.J., Smallwood J., Jackson R.L., Jefferies E. (2024) **Graded and sharp transitions in semantic function in left temporal lobe** *Brain and Language* **251** <https://doi.org/10.1016/j.bandl.2024.105402>
49. Plaut D.C (2002) **Graded modality-specific specialisation in semantics: A computational account of optic aphasia** *Cognitive Neuropsychology* **19**:603–639 <https://doi.org/10.1080/02643290244000112>
50. 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>
51. 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>
52. Trujillo J.P., Holler J (2023) **Interactionally Embedded Gestalt Principles of Multimodal Human Communication** *Perspect Psychol Sci* **18**:1136–1159 <https://doi.org/10.1177/17456916221141422>
53. Hagoort P (2013) **MUC (Memory, Unification Control) and beyond.** *Frontiers in Psychology* **4** <https://doi.org/10.3389/fpsyg.2013.00416>

- 54. 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>
- 55. 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** *Cognitive Neurosci* **21**:2085–2099 <https://doi.org/10.1162/jocn.2008.21161>
- 56. McNeill D. (1992) **Hand and mind : what gestures reveal about thought** University of Chicago Press <https://doi.org/10.2307/1576015>

Author information

Wanying Zhao

CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, Department of Psychology, University of Chinese Academy of Sciences, Beijing, China
ORCID iD: [0000-0002-6979-940X](https://orcid.org/0000-0002-6979-940X)

For correspondence: zhaowy@psych.ac.cn

Zhouyi Li

CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, School of Psychology, Central China Normal University, Wuhan, China

Xiang Li

CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

Yi Du

CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China, Department of Psychology, University of Chinese Academy of Sciences, Beijing, China, CAS Center for Excellence in Brain Science and Intelligence Technology, Shanghai, China, Chinese Institute for Brain Research, Beijing, China 102206
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

Beijing Normal 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.

Weaknesses:

- (1) One major issue is that there is a tight anatomical coupling between pMTG and LIFG. Stimulating one area could therefore also result in stimulation of the other area (see Silvanto and Pascual-Leone, 2008). I therefore think it is very difficult to tease apart the contribution of these areas to the speech-gesture integration process, especially considering that the authors stimulate these regions in time windows that are very close to each other in both time and space (and the disruption might last longer over time).
- (2) Related to this point, it is unclear to me why the HD-TDCS/TMS is delivered in set time windows for each region. How did the authors determine this, and how do the results for TMS compare to their previous work from 2018 and 2023 (which describes a similar dataset+design)? How can they ensure they are only targeting their intended region since they are so anatomically close to each other?
- (3) As the EEG signal is often not normally distributed, I was wondering whether the authors checked the assumptions for their Pearson correlations. The authors could perhaps better choose to model the different variables to see whether MI/entropy could predict the neural responses. How did they correct the many correlational analyses that they have performed?
- (4) The authors use ROIs for their different analyses, but it is unclear why and on the basis of what these regions are defined. Why not consider all channels without making them part of an ROI, by using a method like the one described in my previous comment?
- (5) The authors describe that they have divided their EEG data into a "lower half" and a "higher half" (lines 234-236), based on entropy scores. It is unclear why this is necessary, and I would suggest just using the entropy scores as a continuous measure.

<https://doi.org/10.7554/eLife.99416.1.sa3>

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 the entropy of the stimuli or mutual information, which indicates that the effects stem 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 the 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 the information decoding/integration process (rather than generic excitation/inhibition). The authors' results also showed a 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 fully capture the 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 produces 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 occurs in the brain, and what ERP effects may reflect such a process. 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.

<https://doi.org/10.7554/eLife.99416.1.sa2>

Reviewer #3 (Public review):

In this useful study, Zhao et al. try to extend the evidence for their previously described two-step model of speech-gesture integration in the posterior Middle Temporal Gyrus (pMTG) and Inferior Frontal Gyrus (IFG). They repeat some of their previous experimental paradigms, but this time quantifying Information-Theoretical (IT) metrics of the stimuli in a stroop-like paradigm purported to engage speech-gesture integration. They then correlate these metrics with the disruption of what they claim to be an integration effect observable in reaction times during the tasks following brain stimulation, as well as documenting the ERP components in response to the variability in these metrics.

The integration of multiple methods, like tDCS, TMS, and ERPs to provide converging evidence renders the results solid. However, their interpretation of the results should be

taken with care, as some critical confounds, like difficulty, were not accounted for, and the conceptual link between the IT metrics and what the authors claim they index is tenuous and in need of more evidence. In some cases, the difficulty making this link seems to arise from conceptual equivocation (e.g., their claims regarding 'graded' evidence), whilst in some others it might arise from the usage of unclear wording in the writing of the manuscript (e.g. the sentence 'quantitatively functional mental states defined by a specific parser unified by statistical regularities'). Having said that, the authors' aim is valuable, and addressing these issues would render the work a very useful approach to improve our understanding of integration during semantic processing, being of interest to scientists working in cognitive neuroscience and neuroimaging.

The main hurdle to achieving the aims set by the authors is the presence of the confound of difficulty in their IT metrics. Their measure of entropy, for example, being derived from the distribution of responses of the participants to the stimuli, will tend to be high for words or gestures with multiple competing candidate representations (this is what would presumptively give rise to the diversity of responses in high-entropy items). There is ample evidence implicating IFG and pMTG as key regions of the semantic control network, which is critical during difficult semantic processing when, for example, semantic processing must resolve competition between multiple candidate representations, or when there are increased selection pressures (Jackson et al., 2021). Thus, the authors' interpretation of Mutual Information (MI) as an index of integration is inextricably contaminated with difficulty arising from multiple candidate representations. This casts doubt on the claims of the role of pMTG and IFG as regions carrying out gesture-speech integration as the observed pattern of results could also be interpreted in terms of brain stimulation interrupting the semantic control network's ability to select the best candidate for a given context or respond to more demanding semantic processing.

In terms of conceptual equivocation, the use of the term 'graded' by the authors seems to be different from the usage commonly employed in the semantic cognition literature (e.g., the 'graded hub hypothesis', Rice et al., 2015). The idea of a graded hub in the controlled semantic cognition framework (i.e., the anterior temporal lobe) refers to a progressive degree of abstraction or heteromodal information as you progress through the anatomy of the region (i.e., along the dorsal-to-ventral axis). The authors, on the other hand, seem to refer to 'graded manner' in the context of a correlation of entropy or MI and the change in the difference between Reaction Times (RTs) of semantically congruent vs incongruent gesture-speech. The issue is that the discourse through parts of the introduction and discussion seems to conflate both interpretations, and the ideas in the main text do not correspond to the references they cite. This is not overall very convincing. What is it exactly the authors are arguing about the correlation between RTs and MI indexes? As stated above, their measure of entropy captures the spread of responses, which could also be a measure of item difficulty (more diverse responses imply fewer correct responses, a classic index of difficulty). Capturing the diversity of responses means that items with high entropy scores are also likely to have multiple candidate representations, leading to increased selection pressures. Regions like pMTG and IFG have been widely implicated in difficult semantic processing and increased selection pressures (Jackson et al., 2021). How is this MI correlation evidence of integration that proceeds in a 'graded manner'? The conceptual links between these concepts must be made clearer for the interpretation to be convincing.

<https://doi.org/10.7554/eLife.99416.1.sa1>

Author response:

Responses to Editors:

We appreciate Reviewer 1's first concern regarding the difficulty of disentangling the contributions of tightly-coupled brain regions to the speech-gesture integration process—particularly due to the close temporal and spatial proximity of the stimulation windows and the potential for prolonged disruption. We would like to provide clarification and evidence supporting the validity of our methodology.

Our previous study (Zhao et al., 2021, *J. Neurosci*) employed the same experimental protocol—using inhibitory double-pulse transcranial magnetic stimulation (TMS) over the inferior frontal gyrus (IFG) and posterior middle temporal gyrus (pMTG) in one of eight 40-ms time windows. The findings from that study demonstrated a time-window-selective disruption of the semantic congruency effect (i.e., reaction time costs driven by semantic conflict), with no significant modulation of the gender congruency effect (i.e., reaction time costs due to gender conflict). This result establishes that double-pulse TMS provides sufficient temporal precision to independently target the left IFG and pMTG within these 40-ms windows during gesture-speech integration. Importantly, by comparing the distinctively inhibited time windows for IFG and pMTG, we offered clear evidence of distinct engagement and temporal dynamics between these regions during different stages of gesture-speech semantic processing.

Furthermore, we reviewed prior studies utilizing double-pulse TMS on structurally and functionally connected brain regions to explore neural contributions across timescales as brief as 3–60 ms. These studies, which encompass areas from the tongue and lip areas of the primary motor cortex (M1) to high-level semantic regions such as the pMTG and ATL (Author response table 1), consistently demonstrate the methodological rigor and precision of double-pulse TMS in disentangling the neural dynamics of different regions within these short temporal windows.

Author response table 1.

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

Study	TMS protocol	Results
(1) Teige et al., 2018 <i>Cortex</i>	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 <i>Scientific Reports</i>	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 inter-pulse 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, <i>Journal of Neurophysiology</i>	Double-pulse TMS was applied over left preSMA or left M1 areas at an inter-pulse 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 <i>Cortex</i>	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 <i>Journal of Neuroscience</i>	Double-pulse TMS (dTMS) was delivered over the right posterior superior temporal sulcus (pSTS) 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 pSTS 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 <i>Journal of Cognitive Neuroscience</i>	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 <i>Cortex</i>	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 <i>Journal of Vision</i>	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 <i>Brain Research</i>	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) Kalla et al., 2008 <i>NeuroReport</i>	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 <i>Journal of Neuroscience</i>	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.

Response to Reviewer #1:

(1) For concern on the difficulty of disentangling the contributions of tightly-coupled brain regions to the speech-gesture integration process:

We trust that the explanation provided above has clarified this issue.

(2) For concern on the rationale for delivering HD-tDCS/TMS in set time windows for each region, as well as how these time windows were determined and how the current results compare to our previous studies from 2018 and 2023:

The current study builds on a series of investigations that systematically examined the temporal and spatial dynamics of gesture-speech integration. In our earlier work (Zhao et al.,

2018, J. Neurosci), we demonstrated that interrupting neural activity in the IFG or pMTG using TMS selectively disrupted the semantic congruency effect (reaction time costs due to semantic incongruence), without affecting the gender congruency effect (reaction time costs due to gender incongruence). These findings identified the IFG and pMTG as critical hubs for gesture-speech integration. This informed the brain regions selected for subsequent studies.

In Zhao et al. (2021, J. Neurosci), we employed a double-pulse TMS protocol, delivering stimulation within one of eight 40-ms time windows, to further examine the temporal involvement of the IFG and pMTG. The results revealed time-window-selective disruptions of the semantic congruency effect, confirming the dynamic and temporally staged roles of these regions during gesture-speech integration.

In Zhao et al. (2023, Frontiers in Psychology), we investigated the semantic predictive role of gestures relative to speech by comparing two experimental conditions: (1) gestures preceding speech by a fixed interval of 200 ms, and (2) gestures preceding speech at its semantic identification point. We observed time-window-selective disruptions of the semantic congruency effect in the IFG and pMTG only in the second condition, leading to the conclusion that gestures exert a semantic priming effect on co-occurring speech. These findings underscored the semantic advantage of gesture in facilitating speech integration, further refining our understanding of the temporal and functional interplay between these modalities.

The design of the current study—including the choice of brain regions and time windows—was directly informed by these prior findings. Experiment 1 (HD-tDCS) targeted the entire gesture-speech integration process in the IFG and pMTG to assess whether neural activity in these regions, previously identified as integration hubs, is modulated by changes in informativeness from both modalities (i.e., entropy) and their interactions (mutual information, MI). The results revealed a gradual inhibition of neural activity in both areas as MI increased, evidenced by a negative correlation between MI and the tDCS inhibition effect in both regions. Building on this, Experiments 2 and 3 employed double-pulse TMS and event-related potentials (ERPs) to further assess whether the engaged neural activity was both time-sensitive and staged. These experiments also evaluated the contributions of various sources of information, revealing correlations between information-theoretic metrics and time-locked brain activity, providing insights into the ‘gradual’ nature of gesture-speech integration.

We acknowledge that the rationale for the design of the current study was not fully articulated in the original manuscript. In the revised version, we will provide a more comprehensive and coherent explanation of the logic behind the three experiments, ensuring clear alignment with our previous findings.

| (3) *For concern about the use of Pearson correlation and the normality of EEG data.*

We appreciate the reviewer’s thoughtful consideration. In Figure 5 of the manuscript, we have already included normal distribution curves that illustrate the relationships between the average ERP amplitudes within each ROI or elicited clusters and the three information models. Additionally, multiple comparisons were addressed using FDR correction, as outlined in the manuscript.

To further clarify the data, we will calculate the Shapiro-Wilk test, a widely accepted method for assessing bivariate normality, for both the MI/entropy and averaged ERP data. The corresponding p-values will be provided in the following-up point-to-point responses.

| (4) *For concern about the ROI selection, and the suggestion of using whole-brain electrodes to build models of different variables (MI/entropy) to predict neural responses:*

For the EEG data, we conducted both a traditional region-of-interest (ROI) analysis, with ROIs defined based on a well-established work (Habets et al., 2011), and a cluster-based permutation approach, which utilizes data-driven permutations to enhance robustness and address multiple comparisons. The latter method complements the hypothesis-driven ROI analysis by offering an exploratory, unbiased perspective. Notably, the results from both approaches were consistent, reinforcing the reliability of our findings.

To make the methods more accessible to a broader audience, we will provide a clear description of the methods used and how they relate to each other in the revised manuscript.

Reference:

Habets, B., Kita, S., Shao, Z.S., Ozyurek, A., and Hagoort, P. (2011). The Role of Synchrony and Ambiguity in Speech-Gesture Integration during Comprehension. *J Cognitive Neurosci* 23, 1845-1854. 10.1162/jocn.2010.21462

| (5) *For concern about the median split of the data:*

To identify ERP components or spatiotemporal clusters that demonstrated significant semantic differences, we split each model into higher and lower halves, focusing on indexing information changes reflected by entropy or mutual information (MI). To illustrate the gradual activation process, the identified components and clusters were further analyzed for correlations with each information matrix. Remarkably, consistent results were observed between the ERP components and clusters, providing robust evidence that semantic information conveyed through gestures and speech significantly influenced the amplitude of these components or clusters. Moreover, the semantic information was shown to be highly sensitive, varying in tandem with these amplitude changes.

We acknowledge that the rationale behind this approach may not have been sufficiently clear in the initial manuscript. In our revision, we will ensure a more detailed and precise explanation to enhance the clarity and coherence of this logical framework.

Response to Reviewer #2:

We greatly appreciate Reviewer2's concern regarding whether "mutual information" adequately captures the interplay between the meanings of speech and gesture. We would like to clarify that the materials used in the present study involved gestures performed without actual objects, paired with verbs that precisely describe the corresponding actions. For example, a hammering gesture was paired with the verb "hammer", and a cutting gesture was paired with the verb "cut". In this design, all gestures conveyed redundant meaning relative to the co-occurring speech, creating significant overlap between the information derived from speech alone and that from gesture alone.

We understand the reviewer's concern about cases where gestures and speech may provide complementary rather than redundant information. To address this, we have developed an alternative metric for quantifying information gains contributed by supplementary multisensory cues, which will be explored in a subsequent study. However, for the present study, we believe that the observed overlap in information serves as an indicator of the degree of multisensory convergence, a central focus of our investigation.

Regarding the reviewer's concern about how the neural processes of speech-gesture integration may change with variations in the relative timing between speech and gesture stimuli, we would like to highlight findings from our previous study (Zhao, 2023, *Frontiers in Psychology*). In that study, we explored the semantic predictive role of gestures relative to speech under two conditions: (1) gestures preceding speech by a fixed interval of 200 ms, and (2) gestures preceding speech of its semantic identification point. Interestingly, only in the

second condition did we observe time-window-selective disruptions of the semantic congruency effect in the IFG and pMTG. This led us to conclude that gestures play a semantic priming role for co-occurring speech. Building on this, we designed the present study with gestures preceding speech of its semantic identification point to reflect this semantic priming relationship. Additionally, ongoing research is exploring gesture and speech interactions in natural conversational settings to investigate whether the neural processes identified here are consistent across varying contexts.

To prevent any similar concerns from causing doubt among the audience and to ensure clarity regarding the follow-up study, we will provide a detailed discussion of the two issues in the revised manuscript.

Response to Reviewer #3:

The primary aim of this study is to investigate whether the degree of activity in the established integration hubs, IFG and pMTG, is influenced by the information provided by gesture-speech modalities and/or their interactions. While we provided evidence for the differential involvement of the IFG and pMTG by delineating their dynamic engagement across distinct time windows of gesture-speech integration and associating these patterns with unisensory information and their interaction, we acknowledge that the mechanisms underlying these dynamics remain open to interpretation. Specifically, whether the observed effects stem from difficulties in semantic control processes, as suggested by Reviewer 3, or from resolving information uncertainty, as quantified by entropy, falls outside the scope of the current study. Importantly, we view these two interpretations as complementary rather than mutually exclusive, as both may be contributing factors. Nonetheless, we agree that addressing this question is a compelling avenue for future research. In the revised manuscript, we will include an exploratory analysis to investigate whether the confounding difficulty, stemming from the number of lexical or semantic representations, is limited to high-entropy items. Additionally, we will address and discuss alternative interpretations.

Regarding the concern of conceptual equivocation, we would like to emphasize that this study represents the first attempt to focus on the relationship between information quantity and neural engagement. In our initial presentation, we inadvertently conflated the commonly used term "graded hub," which refers to anatomical distribution, with its usage in the present context. We sincerely apologize for this oversight and are grateful for the reviewer's careful critique. In the revised manuscript, we will clearly articulate the study's objectives, clarify the representations of entropy and mutual information, and accurately describe their association with neural engagement.

Reference

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.

Amemiya, T., Beck, B., Walsh, V., Gomi, H., & Haggard, P. (2017). Visual area V5/hMT+ contributes to perception of tactile motion direction: a TMS study. *Scientific reports*, 7(1), 40937.

Muessgens, D., Thirugnanasambandam, N., Shitara, H., Popa, T., & Hallett, M. (2016). Dissociable roles of preSMA in motor sequence chunking and hand switching—a TMS study. *Journal of Neurophysiology*, 116(6), 2637-2646.

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.

- Pitcher, D. (2014). Facial expression recognition takes longer in the posterior superior temporal sulcus than in the occipital face area. *Journal of Neuroscience*, 34(27), 9173-9177.
- Bardi, L., Kanai, R., Mapelli, D., & Walsh, V. (2012). TMS of the FEF interferes with spatial conflict. *Journal of cognitive neuroscience*, 24(6), 1305-1313.
- D'Ausilio, A., Bufalari, I., Salmas, P., & Fadiga, L. (2012). The role of the motor system in discriminating normal and degraded speech sounds. *Cortex*, 48(7), 882-887.
- Pitcher, D., Duchaine, B., Walsh, V., & Kanwisher, N. (2010). TMS evidence for feedforward and feedback mechanisms of face and body perception. *Journal of Vision*, 10(7), 671-671.
- Gagnon, G., Blanchet, S., Grondin, S., & Schneider, C. (2010). Paired-pulse transcranial magnetic stimulation over the dorsolateral prefrontal cortex interferes with episodic encoding and retrieval for both verbal and non-verbal materials. *Brain Research*, 1344, 148-158.
- Kalla, R., Muggleton, N. G., Juan, C. H., Cowey, A., & Walsh, V. (2008). The timing of the involvement of the frontal eye fields and posterior parietal cortex in visual search. *Neuroreport*, 19(10), 1067-1071.
- Pitcher, D., Garrido, L., Walsh, V., & Duchaine, B. C. (2008). Transcranial magnetic stimulation disrupts the perception and embodiment of facial expressions. *Journal of Neuroscience*, 28(36), 8929-8933.

<https://doi.org/10.7554/eLife.99416.1.sa0>