communications biology

A Nature Portfolio journal

Article

https://doi.org/10.1038/s42003-025-08847-6

Evaluating the temporal order of motor and auditory systems in speech production using intracranial EEG

Siqi Li[1,2,3,11], Zihua Chen[3,4,5,11], Xikang Luo[2,3,11], Jing Wang[6], Pengfei Teng[6], Guoming Luan[6,7], Qian Wang[8,9,10,12] ✉ & Xing Tian[2,3,4,12] ✉

Theories propose that speech production can be approximated as a temporal reversal of speech perception. For example, phonological code is assumed to precede phonetic encoding in the motor system during speech production. However, empirical neural evidence directly testing the temporal order hypothesis remains scarce, mostly because of motor artifacts in non-invasive electrophysiology recordings and the requirements of both temporal and spatial precision. In this study, we investigated the neural dynamics of speech production using stereotactic electroencephalography (sEEG). In both onset latency analysis and representational similarity analysis (RSA), activation in the auditory region of the posterior superior temporal gyrus (pSTG) was observed before articulation, suggesting the availability of auditory phonological code before production. Surprisingly, the activation in the motor region of the inferior frontal gyrus (IFG) preceded that of pSTG, suggesting that the phonological code in the auditory domain may not necessarily be activated before the encoding in the motor domain during speech production.

According to classical theories, speech production is conceptualized as the temporal reversal of speech perception[1,2]. Phonological encoding has been proposed to occur in the auditory-related cortices as an intermediate stage that links language and sensorimotor systems. In speech perception, acoustic signals are processed into higher-level phonological information to support comprehension[3–5]. The posterior superior temporal gyrus (pSTG) is widely considered responsible for phonological encoding in speech perception[6–12]. In speech production, activity is thought to propagate from higher-level language areas to motor-related regions, possibly involving phonological processing in between[1,2]. Some models propose that auditory regions such as the STG are engaged early during speech planning, followed by activation in motor-related areas such as the left posterior inferior frontal gyrus (pIFG), along the dorsal speech pathway[1,13–17]. These models predict that auditory regions would be activated prior to motor-related regions such as Broca's area. However, direct neural evidence comparing the timing of STG and IFG activation during speech production remains scarce.

It is necessary to consider the neural dynamics across major computational units in speech production. Based on neurolinguistic models, activation in the STG is expected around 200–400 ms following stimulus onset. Syllabification is thought to occur between 400–600 ms, primarily engaging the posterior part of the IFG and associated sensorimotor regions involved in transforming phonological units into articulatory commands[1,18,19]. Additionally, phonological processing has been observed in Broca's area, particularly in the left IFG, occurring around 450 ms[20]. However, most evidence regarding the timing of neural activation across these regions comes from behavioral studies, as well as non-invasive electrophysiological and neuroimaging methods. Direct evidence of the timing of activation is crucial for evaluating whether auditory areas consistently activate before motor-related areas during speech production. Intracranial

[1]Key Laboratory of Brain Functional Genomics, East China Normal University, Shanghai, China. [2]Shanghai Key Laboratory of Brain Functional Genomics (Ministry of Education), School of Psychology and Cognitive Science, East China Normal University, Shanghai, China. [3]NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, Shanghai, China. [4]Shanghai Frontiers Science Center of Artificial Intelligence and Deep Learning, Division of Arts and Sciences, New York University Shanghai, Shanghai, China. [5]Neuroscience Institute, New York University Grossman School of Medicine, New York, NY, USA. [6]Department of Neurology, Sanbo Brain Hospital, Capital Medical University, Beijing, China. [7]Beijing Key Laboratory of Epilepsy, Epilepsy Center, Sanbo Brain Hospital, Capital Medical University, Beijing, China. [8]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China. [9]IDG/McGovern Institute for Brain Research, Peking University, Beijing, China. [10]Key Laboratory of General Artificial Intelligence, Peking University, Beijing, China. [11]These authors contributed equally: Siqi Li, Zihua Chen, Xikang Luo. [12]These authors jointly supervised this work: Qian Wang, Xing Tian. ✉e-mail: wangqianpsy@pku.edu.cn; xing.tian@nyu.edu

electroencephalogram (iEEG) offers a promising avenue for directly recording electrophysiological activity with precise temporal and spatial resolutions in the human central nervous system.

To investigate the neural dynamics in the speech production cortical pathway, we conducted stereotactic electroencephalography (sEEG) recordings on patients with medication-resistant epilepsy while they spoke single-character Chinese words. Participants read aloud the written words to avoid any confounds potentially caused by auditory stimuli. The read-aloud process begins in the occipitotemporal cortex, where the initial stage of visual word form processing occurs[21]. This task, combined with sEEG recordings, enabled direct measurement of neural activity across major cortical regions involved in speech production, including the pSTG and IFG.

Results

Behavioral analysis

Participants read aloud single-character Chinese words. Individual articulation reaction times (RTs) were quantified by calculating the time lag between the onset of visual word presentation and the onset of articulation. The average RT was 826.8 ± 106.5 ms. Each individual's RT was used to provide a temporal marker in the following analysis of sEEG data during the production process.

Divergence in activation timing across distinct cortical regions

To investigate neural dynamics across cortical regions during speech production, we first examined the temporal activation patterns of ROIs involved in different stages of the read-aloud task. Onset latency analysis was performed in a representative participant with electrode contacts in both auditory (pSTG) and motor (IFG) regions, enabling within-subject comparison of relative activation timing (Fig. 2A). Specifically, two IFG contacts and six pSTG contacts were included in the analysis (see Supplementary Data 1 for the corresponding MNI coordinates).

We picked a representative participant to evaluate the relative activation timing of the pSTG and IFG during speech production at the subject level. Distinct patterns of neural response power emerged in both the gamma and high-gamma bands (heatmaps, Fig. 2B). When trials were sorted by reaction time (RT), responses in the IFG followed the onset of the visual word stimulus but showed no clear relationship with RT. In contrast, responses in the pSTG occurred much later and were closely aligned with RT (Fig. 2B). The responses from all electrodes of this subject are included in Supplementary Fig. 1, and the averaged responses of all electrode contacts across trials are presented in Fig. 2B, right next to the heatmap for each frequency band.

The responses in the IFG electrodes ($n = 2$) showed latencies of 332.03 ± 2.76 ms (for the single subject, sample average ± std) in the gamma band and 512.70 ± 227.88 ms in the high-gamma band. In contrast, the auditory responses in the pSTG ($n = 6$) were significantly delayed, with latencies of 823.73 ± 69.30 ms in the gamma band and 676.27 ± 168.27 ms in the high-gamma band. Notably, despite the later activation in pSTG, its response still preceded the onset of articulation, occurring earlier than the average reaction time (RT = 839.02 ± 128.26 ms).

We then conduct the onset latency analysis in all participants who had contacts in the regions that mediated the processes of reading aloud, so that the relative activation timing of auditory and motor areas during speech production can be evaluated statistically in the population. Contacts included in the onset latency analysis across various ROIs are visual ROIs (vpIOG: 2, pFG: 26), motor ROI (IFG: 28), and auditory ROIs (HG: 5, pSTG: 26) (Fig. 1C). Patterns emerged in neural response power in both gamma and high gamma bands (heatmaps in Fig. 3A). When arranging single trials according to RT, the responses in the visual ROIs (vpIOG & pFG) did not align with RT, rather that neural activity was immediately induced after the onset of visual word stimuli. Responses in the motor ROI (IFG) followed the responses in the visual ROIs but did not show any clear pattern related to the RT either. Whereas responses in the auditory ROIs (HG & pSTG) were induced much later and were largely in sync with RT.

More quantitative and statistical tests were carried out to test the observed dynamic patterns across the speech production pathway. Neural responses were averaged across trials and the onset latency was determined using a non-parametric temporal cluster method (see details in methods). For the high-gamma band, a rapid response was observed in the electrodes in the visual area vpIOG ($n = 2$) after visual word stimulus presentation, with an onset latency of 71.289 ± 51.758 ms (mean ± SEM). The onset latency of the responses from the electrodes in the pFG ($n = 26$) was 203.576 ± 48.646 ms. The responses in the electrodes in IFG ($n = 28$) had a latency of 368.583 ± 37.081 ms. The onset latency in the

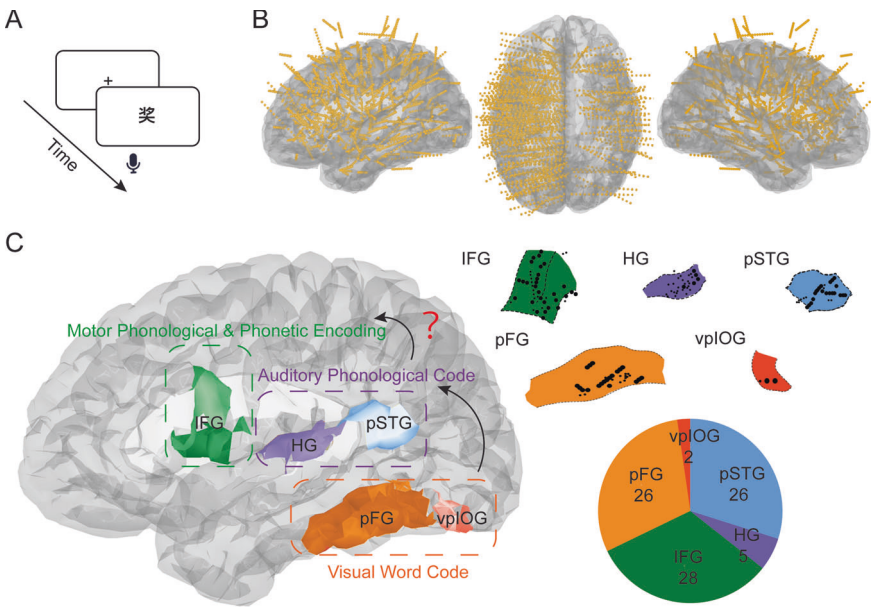

Fig. 1 | Experimental paradigm, sEEG recordings, and hypothesis regarding the processing dynamics of speech production. A Schematic diagram of the experimental paradigm. Following a fixation, a visual single-character Chinese word was presented. Participants were asked to speak out the word as quickly and accurately as possible when the word was displayed on the screen. (Please refer to the Methods for detailed parameters.) **B** The coverage of electrode contacts in all participants. **C** Regions of interest (ROIs) according to the hypothesis about the processing dynamics of speech production. On the left, 5 anatomically defined ROIs are depicted. These ROIs are the areas that mediate three major processes during the read-aloud task, including visual word code (vpIOG & pFG), auditory phonological code (HG & pSTG), and motor-related phonological & phonetic encoding (IFG). On the right, individual contacts are superimposed on the contour of each ROI, demonstrating the distribution of coverage. Smaller dots represent task-related contacts, while larger dots denote the onset-activated contacts. The pie chart summarizes the number of contacts in each ROI that were used in the analyses. The crucial investigation is to assess the temporal order of activation among the ROIs in the auditory and motor systems that mediate the three key processes, as indicated by the arrows.

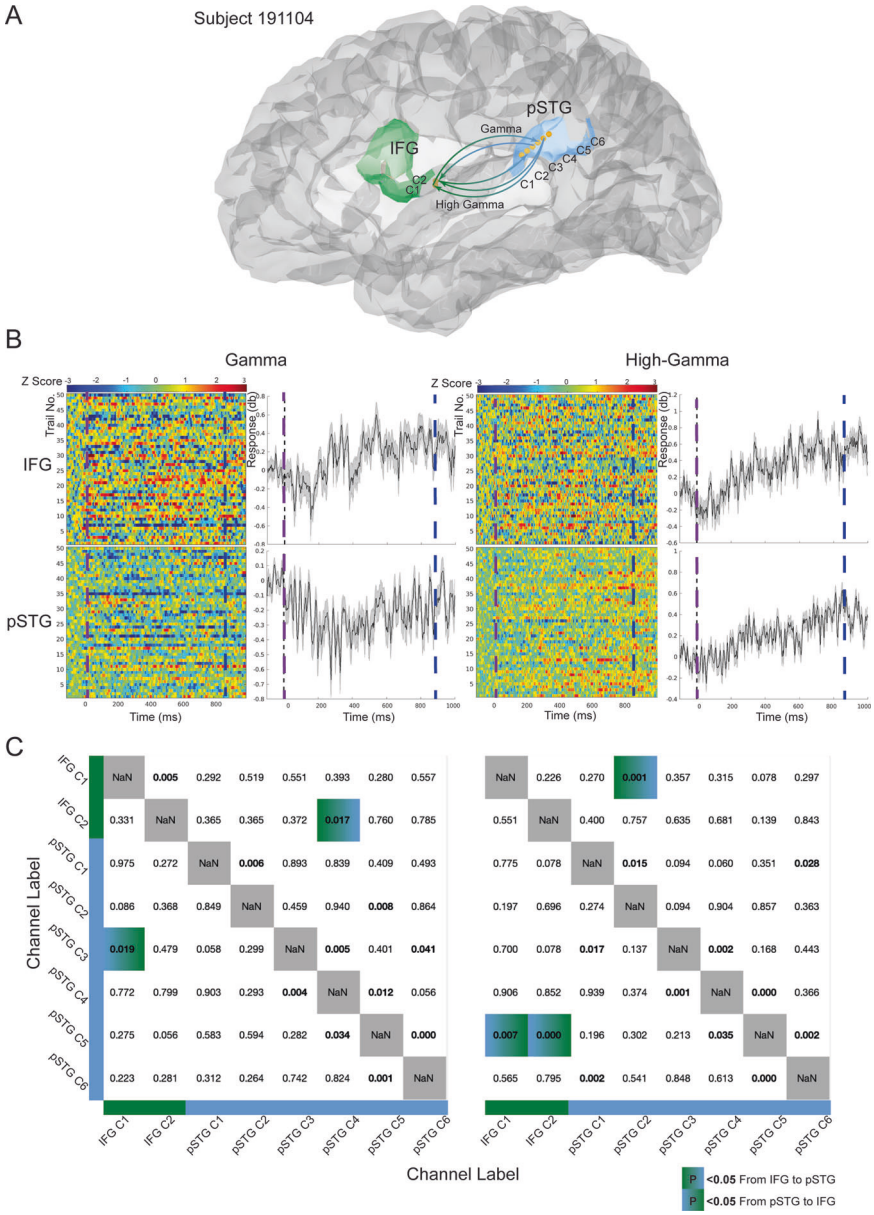

Fig. 2 | Results of activation timing in auditory and motor regions during speech production in a representative participant. A Recordings from contacts in both regions of interest (ROIs) of IFG and pSTG are available in a representative participant (subject 191104). The location of activated contacts with these ROIs is presented on the individual brain atlas, with the number and label of contacts in each ROI that was used in the analyses of this specific subject marked next to each ROI. The arrows represent directional causality results between contacts from the Granger Causality Test (GCT), with details in (**C**). **B** Normalized responses from a representative participant in the IFG and pSTG regions. The left panel shows responses in the gamma band (30–70 Hz), while the right panel shows responses in the high-gamma band (70–140 Hz). For each panel, the left column presents raster plots of averaged responses across contacts within the specified ROI, including all trials sorted by reaction time (error bars denote ± SEM). The right column highlights the averaged response of all contacts within this ROI across trials. Vertical red bars indicate the onset latency of responses in each ROI. **C** Results of the Granger Causality Test (GCT) for Activated Electrode Contact in IFG and pSTG for the same participant. Both axes represent electrode contact labels; two IFG contacts and six pSTG contacts were included in the GCT analysis. The left matrix shows the pairwise *p*-values for the gamma band, and the right one shows results for the high-gamma band. The matrix values indicate the *p*-values from each pairwise test between contacts. Bold numbers indicate that the pair-wise causality result is statistically significant ($P < 0.05$). Green-blue entries denote pairs with statistically significant ($P < 0.05$) causality, either from IFG to pSTG, or from pSTG to IFG.

auditory areas was much later, 617.413 ± 44.706 ms for pSTG ($n = 26$) and 831.250 ± 41.965 ms for HG ($n = 18$). In this case, the activation of pSTG was much earlier than the movement of articulation (RT = 826.8 ± 106.5 ms), supporting that the auditory phonological code is available before production. To examine the temporal sequence of activation between pSTG and IFG, we compared their response latencies. The results showed that pSTG exhibited a significantly longer latency than IFG ($p < 0.001$, two-tailed *t*-test; see Fig. 3B).

For the gamma band, the findings are consistent with those in the high-gamma band. Both the visual areas vpIOG ($n = 2$) and pFG ($n = 5$) exhibit similar response times, with durations of 246.094 ± 238.281 ms and 118.047 ± 6.687 ms, respectively. The IFG region shows a similar pattern to that in the high-gamma band, with an onset latency of 372.638 ± 34.493 ms ($n = 8$). The activation of the auditory cortex had onset latencies of 736.654 ± 78.811 ms for pSTG ($n = 5$) and 764.648 ± 56.343 ms for HG ($n = 5$), both of which were activated before speech onset (RT = 826.809 ± 106.500 ms). Similar to High-gamma band results, the onset latency of pSTG is significantly higher than that of IFG ($p < 0.01$, two-tailed *t* test, bar chart in Fig. 3B). All the onset latency results indicate that the temporal auditory regions activate before the articulation, and the frontal motor-related encoding area activates before the auditory phonological code areas in the pre-articulatory phase of speech production. Similar dynamic response patterns of IFG and pSTG were found when the analyses were aligned to the acoustic onset of speech production (Supplementary Fig. 2).

In addition to the comparison of activation timing, we conducted a Granger Causality Test (GCT) analysis to further investigate the directional functional connectivity between contacts in IFG and pSTG. The analysis was performed on the representative participant who had contact coverage in both regions. The results, as shown in Fig. 2C, revealed frequency-specific patterns of directional connectivity. In the gamma band, we observed significant directional connectivity from IFG to pSTG (e.g., IFG C2 → pSTG C4) as well as from pSTG to IFG (e.g., pSTG C3 → IFG C1). Similarly, in the high-gamma band, we found significant bidirectional connections (e.g., IFG C1 → pSTG C2 and pSTG C5 → IFG C1&2). These results are consistent with the findings in the latency analysis and support dynamic bidirectional interplay between auditory and motor regions during speech production.

Dynamics of representation among cortical areas mediating speech production

The onset latency analysis provided preliminary results of temporal dynamics across cortical areas that mediate hypothetical stages in speech production. Next, we implemented the RSA analysis to further test the processing dynamics at the level of detail representation. We used power spectral density, an acoustic feature of the auditory stimuli, to construct the theoretical RDM that approximated the sensorimotor features in speech production. The empirical RDMs were correlated with the neural responses in the three ROIs, vpIOG, IFG and STG, that were representatives of visual, motor, and auditory processes during read aloud, respectively. Because the theoretical RDM only contained information that was relevant to sensorimotor features but not visual features, we predicted no significant RSA results in the visual ROI. The RSA correlation would be significant in both auditory and motor ROIs; importantly, the timing of the significant correlation results would show distinct dynamic patterns along the time course of speech production.

In the gamma band, the representation was first observed in the IFG around 250 ms for a duration of 50 ms (Fig. 4B). Specifically, in the gamma band, the representation was activated in the IFG at 273–330 ms and in the pSTG at 309–561 ms; whereas in the high gamma band, the same representation was observed at 828–898 ms in the IFG and at 385–486 ms in the pSTG. These results suggest that the auditory-like representation was at least simultaneously induced in both motor and auditory areas – possibly phonetic and phonological representations are activated at a similar time in motor and auditory systems before articulation during speech production. Interestingly, the high-gamma responses in the IFG significantly correlated with the auditory-like representation again right before articulation (onset at 800 ms with a duration around 100 ms). These results hint that distinct frequency band responses could mediate different functions in IFG, where execution-related signals are co-located with and available after the initial encoding of phonetic representation.

When the RSA was applied to the data that were aligned to the acoustic onset of speech production, similar dynamic response patterns were observed (Supplementary Fig. 2). Specifically, in the gamma band, IFG displays an earlier significant cluster compared to pSTG. In the high-gamma band, IFG demonstrates a much later significant cluster that extends continuously up to speech onset, while pSTG shows an earlier, more transient activation cluster.

Discussion

We investigated the neural dynamics of speech production to evaluate whether the phonological code in the auditory area is available before phonetic encoding in the sensorimotor areas involved in speech production. According to the temporal order of speech production, auditory targets should become available immediately after visual word form processing, which predicts that pSTG would be activated following occipitotemporal activity. However, if the processes in speech production do not adhere to this temporal reversal of speech perception, the activation of the auditory-phonological system in pSTG may not necessarily precede the phonetic encoding in IFG. To test this hypothesis, we employed an invasive method with both temporal and spatial precision in human electrophysiological recordings. Our findings reveal that STG was activated before articulation but later than IFG, suggesting that the auditory-phonological code may not necessarily be activated before the encoding in the motor-related system during speech production. Furthermore, IFG activity was observed twice in distinct latencies -- one during preparation and one immediately before articulation, suggesting that the IFG may mediate both motor-related encoding and the execution of speech production. These findings prompt a reevaluation of classical theories and invite a reconsideration of the interaction between auditory and motor systems, especially in the phonological encoding process during speech production.

We primarily focused on the task-related brain regions, including visual input, auditory phonological code, and phonological and phonetic encoding in the motor system. Onset latency analyses revealed distinct activation timing across cortical regions, highlighting task-specific patterns. The fusiform gyrus plays a role in linking visual word forms with speech sounds[22]. In this study, we observed neural responses that were immediately time-locked to the visual word onset in visual cortical areas of vpIOG and pFG. Previous studies have shown that visual word identification occurs within 170 to 250 ms[1,23,24], aligning with the approximately latency of 200 ms in the visual areas that we observed, which demonstrates the validity of the latency analysis and its results in intracranial recordings.

Next, in the IFG, presumably an area for motor-related encoding, the onset latencies of neural activity started at 200 ms in both gamma and high-gamma frequency bands. However, auditory areas, including HG and pSTG that presumably mediate the retrieval of auditory phonological code, activated last with onset latencies close to but before the time of articulation. First, the observation of auditory cortices activation before articulation is, to our knowledge, the first clear and direct neural evidence suggesting the availability of auditory phonological code before production. Moreover, such 'IFG-first, STG-second' temporal order in speech production is consistent with findings of early motor activity in several recent intracranial studies, but inconsistent with the serial processing hypothesis[1]. Syllabification and other phonological processes have been proposed as occurring between 400 to 600 ms after phonological code retrieval at around 200 to 400 ms[1]. However, evidence from intracranial recordings shows that Broca's area is activated around 200 ms for word identification and around 450 ms for phonological processing[20]. Similarly, Broca's area activates during the pronunciation encoding phase approximately 250 ms after stimulus presentation[14]. These intracranial recording results are consistent with our observations that IFG activation precedes that of the pSTG (Fig. 3), suggesting that motor-related encoding may begin before the auditory phonological code is completely available; the later activation in auditory cortices may reflect feedback processes related to speech production control.

While this observed timing order of IFG preceding pSTG aligns with the motor-first prediction, the ~200–300 ms latency difference alone does not provide decisive support for a strictly serial processing model. Both regions were activated well before articulation, which leaves open the possibility of interactive or parallel processing between motor and auditory systems. While our findings challenge classical serial models of speech production, they do not fully overturn them. Instead, these results may be more consistent with frameworks emphasizing dynamic, bidirectional, or predictive interactions. Additional data and future studies are needed to clarify the precise nature of these processes.

That IFG precedes STG may manifest a more efficient process in speech production and control. A motor control theory, called internal forward models[25–27], proposes that after the programmed motor commands, a copy of motor signals is transmitted to sensory regions and serves as a predictive signal[28,29]. In speech production, this copy of motor signal may originate in the IFG and is transformed into a predicted auditory signal in the pSTG for controlling an early stage of speech production[30–32]. Prior research suggests that efference copies play a crucial role in sensorimotor integration by predicting the sensory consequences of forthcoming articulatory movements. Notably, efference copies have been shown to suppress auditory cortical responses specifically within the high-gamma band during speech production, highlighting the contribution of gamma-band activity to internal prediction mechanisms. Moreover, gamma oscillations have been closely linked to both feedforward sensorimotor signaling and predictive coding. Our connectivity analysis revealed bidirectional functional connectivity between IFG and pSTG (Fig. 2C), which is consistent with the hypothesis of efference copy and internal forward model in speech production control. Taken together, these findings support the view that gamma activity in the IFG likely reflects an integration of motor-related encoding and predictive signaling, rather than a single, isolated process. Another possibility is that this motor-based auditory prediction may enhance content-specific auditory goals[33], and hence facilitate speech preparation. Regardless, the observation of IFG preceding STG indicates a tightly coupled interaction, where the motor-related encoding in IFG may occur first and generate prediction and interact (compare or modulate) with

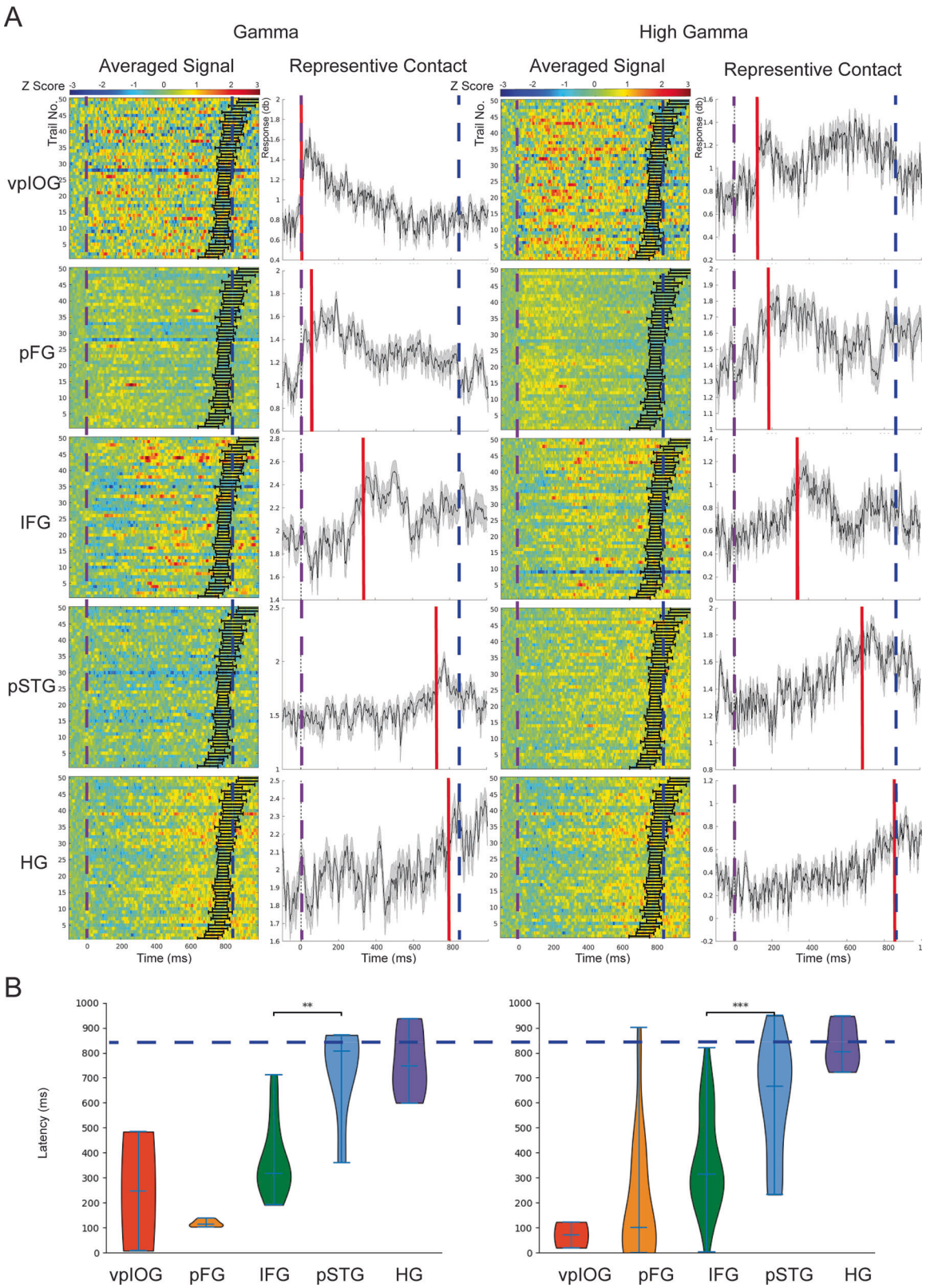

Fig. 3 | Results of latency analysis across all participants. A Normalized responses for all participants in the ROIs. The left panel shows responses in the gamma band (30–70 Hz), while the right panel shows responses in the high-gamma band (70–140 Hz). For each panel, the left column presents raster plots of averaged responses across contacts within a given ROI, with trials sorted by reaction time (error bars denote ± SEM). The right column highlights the averaged response of a representative contact across trials. Vertical red bars indicate the onset latency of responses in each ROI. Vertical violet dashed lines mark the onset of the visual word stimulus, and vertical blue dashed lines indicate the average reaction time. **B** Comparison of onset latency between ROIs. The onset latency of pSTG is significantly longer than that of IFG, in both Gamma ($^{**}p < 0.01$, two-tailed t-test) and High-Gamma ($^{***}p < 0.001$, two-tailed t-test) bands. The horizontal blue dashed lines indicate the average reaction time.

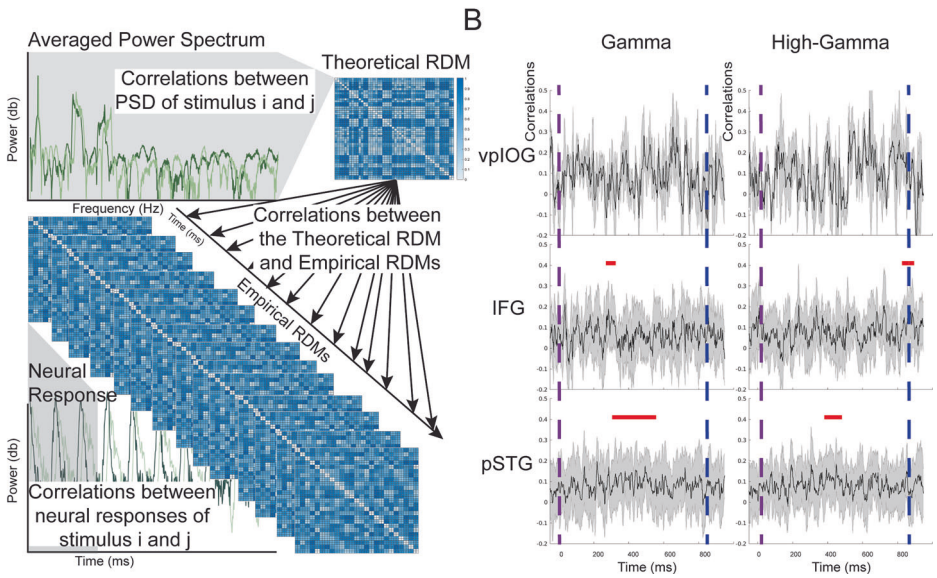

Fig. 4 | Results of temporal-progressive RSA analysis. A Procedures of the temporal-progressive RSA analysis. Theoretical RDM was constructed by comparing the averaged acoustic PSD of a given pair of stimulus words. sEEG data were epoched and averaged with a window of 25 time points (~50 ms) and each empirical RDM was created by comparing the time-averaged responses of producing a given pair of stimuli. The correlation between theoretical and empirical RDMs was obtained for each of the time windows. **B** RSA results, separately in gamma and high-gamma bands. The black line represents the mean correlation between RDMs averaged across contacts in a given ROI. The shaded area indicates the standard deviation. The red horizontal line marks the significant positive clusters (cluster-based non-parametric test, $p < 0.05$). The vertical violet and blue dashed lines indicate the stimulus onset and average reaction time, respectively.

the auditory target in STG. Such interactive processes are crucial in feedback integration and dynamic speech production control.

The GCT results provide valuable insights regarding the bidirectional functional connectivity between motor and auditory regions during speech production (Fig. 2C). Noted that electrode placement is clinically determined and varies across subjects, which constrains the scope of investigation for providing a holistic picture of processes in these regions. Nevertheless, a degree of local consistency was observed, particularly the robust bidirectional connectivity between IFG and pSTG contacts 4, 5, and 6, which indicates the reliability of our findings. The connectivity results are consistent with the results in latency analyses, highlighting a flexible and reciprocal mechanism between motor and auditory regions that mediates speech production.

RSA results replicated the latency results of temporal activation order in the motor and auditory systems and revealed a more complex dynamics, providing additional evidence on activation timing and sequence from a representational perspective. In the gamma band, distinct speech-related information was first decoded in the IFG, followed by decoding in the pSTG. Although the onset latency of pSTG in the RSA results was earlier than that observed in the latency analysis, the relative temporal order of the motor and auditory activation was the same, providing consistent results suggesting that motor-related encoding may initiate before the auditory phonological code. The latency differences of STG in RSA and latency analysis may indicate interaction between motor and auditory systems during speech production control in a hierarchical manner. The observation of earlier STG activation in RSA analysis may be the result of the immediately preceding activation in IFG via the internal forward model. Whereas the later STG activation observed in the latency analysis was immediately before the articulation. The later STG activation may be the result of the motor system activation, presumably the primary motor cortex or premotor cortex, which is close to the time of articulation in the later stage of speech production.

The RSA results also revealed IFG activation close to the time of articulation, but in the high-gamma band (Fig. 4B). These temporal characteristics indicate that IFG may activate again when close to the articulation. That is, after the control during preparation, the IFG that is in the upstream of speech production control may activate again and send to the downstream (presumably the premotor and primary motor cortices) to execute. Moreover, the representation of the later IFG activation was in a different frequency band than the earlier IFG activation that was observed in a lower range of the gamma band.

Neural activities in both the gamma and high-gamma frequency bands play pivotal roles in neural communication and information processing, but they reflect distinct aspects of cortical dynamics. Gamma activity, thought to arise from rhythmic interactions between reciprocally connected inhibitory interneurons and excitatory pyramidal neurons, primarily reflects input to cortical circuits and intracortical processing within excitatory-inhibitory feedback loops[34]. Gamma activity is strongly correlated with the BOLD signal measured by fMRI[35] and demonstrates selective sensitivity to various stimulus features, providing a mechanism for routing and prioritizing information during task performance[36,37]. Additionally, gamma oscillations support the coordination of activity across distributed neural networks, playing a crucial role in effective neural communication[38,39].

In contrast, high-gamma activity, associated with local neural processes, reflects highly active and synchronized neuronal populations engaged in various functions[40–44]. It is a broadband, non-oscillatory signal likely arising from the summed postsynaptic potentials of large neuronal populations[34,45–47], and is traditionally regarded as a proxy for local ensemble spiking activity[45,48–50]. High-gamma power is more directly linked to local neuronal activity, particularly involving synchronized populations during motor execution. This distinction aligns with electrophysiological findings suggesting that gamma activity relates to broader network-level computations, whereas high-gamma activity more closely reflects local spiking and neuronal synchrony[48]. Furthermore, high-gamma activity has been implicated in task encoding and the integration and transmission of information across neural circuits[51].

Taken together, gamma and high-gamma frequency bands may reflect distinct representations and computations within different neural populations and microcircuits. Accordingly, the activity we observed in the same region of the IFG but across different time windows and frequency bands may collectively represent control processes at the encoding and execution stages of speech production, mediated by specific neural subpopulations or microcircuits within the same area. Our findings suggest that early gamma activity in the IFG likely reflects a combination of motor-related encoding and predictive processes, whereas later high-gamma activity corresponds more directly to movement execution. Some methodological overlap between these processes may exist, which future work could help to resolve.

Our novel results indicate a potential alternative to the serial temporal processing model of speech production. Current consensus views speech production as a temporal reversal of speech perception – sharing the same phonological code and transforming the phonological code from the auditory to motor domain[52–54]. However, direct neural evidence supporting the strict temporal order of 'auditory phonological code first then motor-related encoding' is limited. Several models implicate that the STG mediates the process of phonological code retrieval during speech production[1,2].

However, using sEEG recordings to avoid motor artifacts and meet the rigorous requirement of temporal and spatial precision, our onset latency results showed that STG activation after IFG, which is consistent with an alternative account that motor areas, including IFG, may be the areas translating the phonological codes in overt speech. Additionally, our RSA analysis indicates that both the IFG and pSTG encode similar representations, but again, the IFG had an earlier latency than that of pSTG, which further supports the alternative view. These results collectively hint that the auditory phonological code may not always precede motor-related encoding during speech production.

This study provides preliminary insights that may inform future research on the temporal coordination of auditory and syllabification processes during speech production. First, by combining onset latency analysis, RSA, and GCT within the same sEEG data set, this study offers preliminary evidence consistent with an earlier-than-expected involvement of the IFG in speech preparation, potentially preceding auditory cortical activation. These results may be compatible with interactive or predictive models of speech production, but further validation is required. The high temporal and spatial resolution of sEEG in this study offers a unique advantage in investigating the neural dynamics of speech production, especially where articulatory artifacts limit the investigation using non-invasive scalp recordings. Second, bidirectional IFG–pSTG interactions in both the gamma and high-gamma bands may reflect neural dynamics consistent with internal forward models —mechanisms that have long been theorized in speech production but are difficult to capture at a fine scale in both temporal and spatial domains.

Nonetheless, caution is warranted. Because the present evidence derives from a single-participant GCT analysis and a relatively small set of Chinese syllables, the proposed timing hierarchy and connectivity pattern should be re-examined in group-level studies that sample broader phonological inventories and, where feasible, combine invasive with non-invasive recordings. Such data would make it possible to replicate the earlier IFG engagement, model articulatory–phonetic dimensions explicitly in the RDMs, and disentangle predictive from feedback-driven signals with greater precision. The methodological framework and preliminary results in the current study will inspire future work for investigating these questions in larger populations.

Several other limitations should be noted. First, as a well-recognized limitation of sEEG studies due to the clinical implantation of electrodes, we have limited coverage in some areas of interest, especially the visual areas like vpIOG, which might influence the robustness of onset latency in these specific regions. As a result, we remain cautious not to overinterpret the results from less-well-covered areas. Moreover, the data were acquired from patients with drug-resistant epilepsy, which presents general limitations related to the clinical nature of the population. Although both epileptic pathology and antiepileptic medication may influence neural signals, we carefully excluded electrodes within seizure onset zones and included only those in non-pathological tissue, distant from any ictal or interictal discharges. Given the broad cortical coverage, preserved behavioral performance, and the fact that our conclusions rely on relative timing between regions, we believe the observed neural dynamics reflect genuine cognitive processes rather than artifacts of epilepsy or its treatment.

The presented RSA approach, though novel in sEEG research, also brings limitations in inferring the dynamics in representational coding and transformation in the distributed neural system. Specifically, RSA inherently requires aggregating information across electrode contacts and time windows to ensure sufficient statistical power. This aggregation, while necessary, may obscure fine-grained spatial and temporal dynamics of neural representations, particularly how they evolve and propagate across time and regions. Moreover, the current study used power spectral density (PSD) to construct the theoretical RDM, which primarily captures low-level acoustic similarities. Albeit informative, PSD does not fully reflect higher-order phonological abstractions. As such, the representational inferences from RSA should be interpreted with caution, particularly regarding the extent to which these patterns reflect phonological—as opposed to acoustic—processing. Future studies could benefit from using more linguistically informed features, such as phonological, phonetic or articulatory characteristics, to construct RDMs that better capture the complexity of speech representations. Furthermore, the current experimental design, which includes 50 distinct sounds and accommodates the phonological complexity of Chinese pronunciation, does not offer sufficient statistical power to reliably model detailed phonetic and phonological features (like place and manner of articulation) for the RDM. We are conducting follow-up experiments that use a smaller selected set of stimuli to model more precise phonetic and phonological features. Additionally, taking individual variability into account -- such as recording each subject's pronunciations and constructing single-trial RDMs -- could offer stronger evidence for the neural mechanisms underlying speech production.

In conclusion, using high temporal and spatial resolution sEEG recordings in an articulation task, we found that before the movement of speech production, the STG was activated, suggesting the phonological code was encoded in an auditory form before articulation. Additionally, we observed that IFG activation preceded STG, suggesting that the motor system may begin programming speech codes even before the complete retrieval of an auditory target. These findings provide evidence of neural dynamics for the timing and order of processing stages in the speech production network, suggesting that auditory phonological encoding does not always precede motor-related processes. The human intracranial electrophysiological study of speech production highlights the importance of neural circuitry level evidence for evaluating psycholinguistic theories and encourages a thorough reconsideration of functional dynamics between language and motor systems during speech production.

Materials and methods

Participants

A total of 26 patients (16 males, age ranged from 14 to 46 years) with drug-resistant epilepsy, who underwent invasive stereo-electroencephalogram monitoring for potential surgical interventions at the Sanbo Brain Hospital of Capital Medical University (Beijing, China) participated in this experiment. All participants were native Mandarin speakers, right-handed and self-reported that hearing, vision, or corrected vision were normal. Written informed consent was obtained from every participant before the experiment. The experimental protocol was approved by the Ethics Committee at the Sanbo Brain Hospital of Capital Medical University and New York University Shanghai. All ethical regulations relevant to human research participants were followed.

Experimental procedures

The experimental materials comprised 50 single-character Chinese words selected based on their relatively high frequency of occurrence (on average 80.05 times per million)[55]. Visual stimuli were presented on an LCD display, and participants' vocalizations were captured using a professional microphone (SHURE Beta 58 A). Each of the 50 Chinese words was presented in a random order across 50 trials. A fixation cross was displayed for 500 ms, followed by the presentation of the Chinese word stimuli. The visual stimulus was presented in white at the center of the screen for a duration of 1100 to 1600 ms with random jitter; the stimulus remained visible even as articulation started. Participants were instructed to speak each word as fast and accurately as possible within a 3 s time frame, starting from the onset of the visual word stimulus. The onset time of vocal responses was recorded to determine reaction times (RTs), which are defined by the measured as the time duration between the onset of the visual cue and the onset of articulation. The experimental paradigm is illustrated in Fig. 1A.

Recording

All 26 patients were implanted with stereo-electrodes, in total of 2916 electrode contacts (Fig. 1B). Each electrode had 8-16 independent recording contacts, measuring 0.8 mm in diameter, 2 mm in length, spaced 3.5 mm apart (Huake Hengsheng Medical Technology). Implantation sites for the EEG electrodes were determined solely based on clinical criteria. Intracranial EEG signals were sampled at 512 Hz using the Nicolet video-EEG

monitoring system (Thermo Nicolet Corp., USA) without any online filtering. The signal from each recording contact was amplified using a Nicolet clinical amplifier. All signals were referenced online to a forehead scalp electrode, following the default configuration of the clinical recording system. Subsequent data processing was performed offline.

Data analysis

Behavioral data analysis. The reaction time (RTs) of the articulation were calculated as the time lag between the onset of the visual stimuli and the onset of the vocalization.

Electrode localization. To localize the sites of stereo-electrodes, we combined the anatomical information obtained from preoperative magnetic resonance imaging (MRI) with the positional data of the electrodes from postoperative computer tomography (CT). For each patient, post-implantation CT images were co-registered with pre-implantation T1-weighted MRI scans using BioImage software (http://bioimagesuite.yale.edu)[56]. Individual electrodes were identified on the aligned CT images, and the coordinates of electrode contacts were obtained using the Brainstorm toolbox (http://neuroimage.usc.edu/brainstorm)[57]. Cortical surfaces were reconstructed from preoperative MRI data using BrainSuite software (http://brainsuite.org/)[58]. To assign anatomical labels to each contact, subcortical and cortical segmentations were conducted based on individual preoperative T1 MRI scans using the individual atlas (USCBrain)[59]. Region of interest (ROI) labels for each channel were visually inspected and manually corrected as needed. Contact coordinates were normalized to the MNI space.

Preprocessing. The intracranial signals underwent preprocessing using the EEGLAB toolbox[60]. For each electrode contact, we extracted the signals in the gamma band (30–70 Hz) and high-gamma band (70–140 Hz) from the continuous sEEG data. The power envelope of the bandpass-filtered gamma and high-gamma signals was derived using the Hilbert transform. Subsequently, the filtered dataset was segmented into epochs spanning from −100 to 1000 ms relative to the onset of the visual word stimulus, with baseline correction applied using the -100 ms pre-stimulus period. Epochs containing recording artifacts or interictal epileptic spikes were identified through visual inspection and excluded from further analysis. The remaining epochs were utilized for subsequent analyses. To apply universal subject analysis, all signals were normalized. Specifically, the mean μ and variance σ^2 of all time points during the baseline were determined initially. Thereafter, normalized signals were computed using the z-score method as Eq. 1, where V symbolizes the original signal, μ represents the signal sample mean, and σ is for sample std.

$$\text{normalized signal} = \frac{V - \mu}{\sigma^2} \quad (1)$$

For the intracranial signals that were re-aligned to the acoustic onset of speech production, data were normalized according to the duration between stimulus onset and reaction onset across trials and subjects by linearly upsampling or downsampling each segment between stimulus onset and recorded acoustic onset of each trial to a uniform length of 500 time points[61].

Task-related electrodes. Electrode contacts afflicted by sustained artifacts and seizure spikes were excluded. The selection of task-related electrodes was determined by the existence of a significant deviation between the post-stimulus signal and the preceding baseline signal. The statistical test employed the Bootstrap method[14,62,63]. The disparity between average signals before and after stimulus presentation was calculated, followed by a shuffle of all time points. The difference was recalculated after each shuffle, with the distribution from 2000 shuffles serving as the null hypothesis distribution. The empirical difference between the baseline and post-stimulus signals was integrated into the null hypothesis distribution to test if the empirical difference was statistically significant ($p < 0.05$). Electrode contacts that showed significant differences between the baseline and post-stimulus signals were designated as task-related electrodes. The task-related electrodes from all subjects were further specified in five regions of interest (ROIs) to extract more robust regional trends without overinterpreting the variability of individual electrode contacts. The contacts localized in Heschl's gyrus (HG) and posterior superior temporal gyrus (pSTG) were selected from the auditory region. Contacts in the ventroposterior inferior occipital gyrus (vpIOG) and posterior fusiform gyrus (pFG) were designated as from the visual-related region, while those in the pars opercularis (IFG) were designated as from the motor-related region (Fig. 1C).

Based on task-related electrodes, the subsequent onset latency and RSA analysis focused on neural activity in both the gamma and high-gamma bands due to their significance in understanding neural communication and information processing. Gamma activity, which reflects the behavior of cortical networks where populations of excitatory and inhibitory neurons interact and occurs at latencies consistent with task performance timing[36,37], and high-gamma activity, indicating highly synchronized neuron populations, were both analyzed to capture the dynamics of task-specific activations[41–44].

Onset latency analysis. Onset latency analysis was applied to all task-related electrode contacts as a group analysis. The response onset is computed on a cross-trial, by-electrode basis using the re-sampling approach[14,62,63]. After obtaining the normalized gamma and high-gamma responses, the null hypothesis distribution is estimated by resampling the average signal of the baseline period 2000 times. The signal subsequent to stimulus presentation was integrated into the null hypothesis distribution, adopting a significance threshold of $\alpha = 0.05$. When 25 consecutive time points surpass this threshold, the leftmost time point of the cluster is recognized as the response onset. To examine the temporal sequence of activation between ROIs, their response latencies were compared using a two-tailed t-test. Additionally, to ensure consistent temporal alignment across trials, all latency analyses were time-locked to the onset of the visual cue, which served as a reliable reference point and mitigated potential misalignment between articulatory movement and acoustic onset.

Granger causality test (GCT) analysis. To assess directional functional connectivity between the inferior frontal gyrus (IFG) and posterior superior temporal gyrus (pSTG), we performed a Granger Causality Test (GCT) analysis[64–66] on a representative subject (191104, contact location shown in Fig. 2A, contact coordinate in MNI space available in Supplementary Data 1), with two activated electrode contacts in the IFG and six activated contacts in the pSTG (contact location shown in Fig. 2A). For each channel in both the gamma and high-gamma bands, multivariate time series were modeled using a vector autoregressive (VAR) model, with model order selected via the Akaike Information Criterion (AIC) and parameters estimated using the Nuttall-Strand algorithm. Granger causality tests (GCT) were performed by comparing a full VAR model—including all predictor variables—to a reduced model in which the putative causal source channel was excluded from the prediction of the target channel. The test statistic was computed as the log-likelihood ratio of the residual covariances between these models. Under the null hypothesis of no Granger causality, this statistic follows a chi-square distribution, from which p-values were derived to determine statistical significance. P-values from the GCT were computed for each contact pair and organized into a matrix, with electrode labels represented along both axes (Fig. 2C). Two distinct color-combination boxes highlight the different directional connectivity in inter-region contact pairs that reach statistical significance ($p < 0.05$, uncorrected). Separate analyses were conducted and matrices were generated for the gamma and high-gamma bands to explore potential frequency-specific patterns of the directional connectivity.

Representation similarity analysis (RSA). The core workflow of RSA involves selecting representations, constructing empirical and theoretical Representation Dissimilarity Matrices (RDMs), and comparing them[67]. For this study, we constructed the theoretical RDM from the power spectral density (PSD) of the acoustic words.

To further minimize variability in acoustic properties, we synthesized the sound of all 50 Chinese monosyllabic words used in the experiment with a Chinese text-to-speech toolkit[68]. We extracted the PSD matrix for each word using the discrete Fourier transform spectrogram algorithm, spanning a range of 0.1–5000 Hz with a window length of 260 and a 0.1 Hz increment[69,70]. Correlations between PSD matrices of all possible pairs of words were obtained. The dissimilarity was determined by subtracting the correlation value from 1, yielding the theoretical RDM. We employed similar procedures to construct the empirical neural RDMs. sEEG data were aligned to the stimulus onset, and all channels from the same ROI across all subjects were pooled together. Data from each ROI were then epoched using a window comprising 25 time points, and the window was shifted in a step-wise manner of a single time point, ensuring a balance between data diversity and statistical power. The choice of 25 time points (about a 50 ms time window in the 512 Hz sampling rate) aligns with the minimum duration required to process constituent phonetic segments[71]. For each time window of every channel, correlations between the sEEG responses to all possible pairs of words were obtained. The dissimilarity was computed by subtracting the correlation value from 1, yielding the empirical RDMs as a function of time.

Subsequently, we compared the similarity between the theoretical and empirical RDMs using a non-parametric temporal-cluster-based permutation test[63]. Specifically, the correlation between the theoretical RDM and empirical RDM at a given time was obtained by conducting Pearson correlation across all contacts in an ROI. The correlation values were subject to a *t*-test against zeros. The adjacent data points that exceeded the threshold ($p < 0.05$) were grouped to form temporal clusters, and the cluster-level empirical statistics were obtained by summing the *t*-value of all time points in a cluster. Randomly shuffling the group/condition labels of the data and repeating the steps above for 2^n times (n for the number of electrode contacts in ROI) yielded a distribution of cluster-level statistics under the null hypothesis distribution. The empirical cluster-level statistics that exceeded a value at 95% of the null distribution were considered statistically significant. The RSA analysis workflow is shown in Fig. 4A.

Statistics and reproducibility. All statistics were done in MATLAB[72] (version R2020b; MathWorks, Natick, MA, USA) environment, and a universal significance threshold of 0.05 was applied across all statistical tests. Task-related electrodes were identified using a bootstrap approach, with 2000 shuffles generated to construct the null hypothesis distribution. Onset latency analysis involved 200 resampling iterations, followed by a two-tailed *t*-test. For Granger causality (GCT) analysis, test statistics were calculated as the log-likelihood ratio of residual covariances between competing models, and the uncorrected *p*-values were derived from a chi-square distribution. *P*-values for representational similarity analysis (RSA) were obtained through cluster-based permutation pooling.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data from the representative participant of all experimental conditions and all supplementary data tables are included in the GitHub repository (https://github.com/zc2214/Speech_sEEG) for verifying reproducibility and ensuring transparency. Because of ethical restrictions, the full dataset cannot be publicly archived. Readers interested in accessing the full dataset may contact the lead author. Data access will be granted to named individuals in accordance with ethical guidelines for the reuse of clinical data, subject to the completion of a formal data-sharing agreement and institutional approval.

Code availability

Custom MATLAB scripts and functions used for behavioral data analysis, preprocessing, electrode localization, statistical analyses, and Representation Similarity Analysis (RSA) are publicly available at https://github.com/zc2214/Speech_sEEG. These scripts were developed using MATLAB[72] (version R2020b; MathWorks, Natick, MA, USA) and are compatible with the open-source toolboxes employed in this study, including EEGLAB[60] (version 14.1.2), Brainstorm[57] (version 3.5), and BrainSuite[58] (version 19a).

Received: 11 December 2024; Accepted: 4 September 2025;
Published online: 08 October 2025

References

1. Indefrey, P. & Levelt, W. J. M. The spatial and temporal signatures of word production components. *Cognition* **92**, 101–144 (2004).
2. Levelt, W. J. M., Roelofs, A. & Meyer, A. S. A theory of lexical access in speech production. *Behav. Brain Sci.* **22**, 1–38 (1999).
3. Price, C. J. The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann. N. Y. Acad. Sci.* **1191**, 62–88 (2010).
4. Scott, S. K. & Johnsrude, I. S. The neuroanatomical and functional organization of speech perception. *Trends Neurosci.* **26**, 100–107 (2003).
5. Scott, S. K. & Wise, R. J. The functional neuroanatomy of prelexical processing in speech perception. *Cognition* **92**, 13–45 (2004).
6. Benson, R. R. et al. Parametrically dissociating speech and nonspeech perception in the brain using fMRI. *Brain Lang.* **78**, 364–396 (2001).
7. Buchsbaum, B. R., Hickok, G. & Humphries, C. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. *Cogn. Sci.* **25**, 663–678 (2001).
8. Chang, E. F. et al. Categorical speech representation in human superior temporal gyrus. *Nat. Neurosci.* **13**, 1428–1432 (2010).
9. Hamilton, L. S., Edwards, E. & Chang, E. F. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Curr. Biol.* **28**, 1860–1871.e4 (2018).
10. Humphries, C., Sabri, M., Lewis, K. & Liebenthal, E. Hierarchical organization of speech perception in human auditory cortex. *Front. Neurosci.* **8**, 111344 (2014).
11. Yang, Z. & Long, M. A. Convergent vocal representations in parrot and human forebrain motor networks. *Nature* **640**, 427–434 (2025).
12. Castellucci, G. A. et al. Neural activity flows through cortical subnetworks during speech production. *bioRxiv* 2025–06 (2025).
13. Bohland, J. W., Bullock, D. & Guenther, F. H. Neural representations and mechanisms for the performance of simple speech sequences. *J. Cogn. Neurosci.* **22**, 1504–1529 (2010).
14. Flinker, A. et al. Redefining the role of Broca's area in speech. *Proc. Natl. Acad. Sci. USA* **112**, 2871–2875 (2015).
15. Hickok, G. Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* **13**, 135–145 (2012).
16. Wilson, S. M., Saygin, A. P., Sereno, M. I. & Iacoboni, M. Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* **7**, 701–702 (2004).
17. Tian, X. & Poeppel, D. The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *J. Cogn. Neurosci.* **25**, 1020–1036 (2013).
18. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**, 393–402 (2007).
19. Indefrey, P. The spatial and temporal signatures of word production components: a critical update. *Front. Psychol.* **2**, 255 (2011).
20. Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D. & Halgren, E. Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* **326**, 445–449 (2009).
21. Dehaene, S. & Cohen, L. The unique role of the visual word form area in reading. *Trends Cogn. Sci.* **15**, 254–262 (2011).
22. Striem-Amit, E., Cohen, L., Dehaene, S. & Amedi, A. Reading with sounds: sensory substitution selectively activates the visual word form area in the blind. *Neuron* **76**, 640–652 (2012).

23. Gaillard, R. et al. Direct intracranial, FMRI, and lesion evidence for the causal role of left inferotemporal cortex in reading. *Neuron* **50**, 191–204 (2006).
24. Marinkovic, K. et al. Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron* **38**, 487–497 (2003).
25. Kawato, M. Internal models for motor control and trajectory planning. *Curr. Opin. Neurobiol.* **9**, 718–727 (1999).
26. Schubotz, R. I. Prediction of external events with our motor system: towards a new framework. *Trends Cogn. Sci.* **11**, 211–218 (2007).
27. Wolpert, D. M. & Ghahramani, Z. Computational principles of movement neuroscience. *Nat. Neurosci.* **3**, 1212–1217 (2000).
28. Sperry, R. W. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J. Comp. Physiol. Psychol.* **43**, 482 (1950).
29. von Holst, E. & Mittelstaedt, H. Das Reafferenzprinzip. *Naturwissenschaften* **37**, 464–476 (1950).
30. Chu, Q., Ma, O., Hang, Y. & Tian, X. Dual-stream cortical pathways mediate sensory prediction. *Cereb. Cortex* **33**, 8890–8903 (2023).
31. Li, Y., Luo, H. & Tian, X. Mental operations in rhythm: Motor-to-sensory transformation mediates imagined singing. *PLoS Biol.* **18**, e3000504 (2020).
32. Zhang, W., Yang, F. & Tian, X. Functional connectivity between parietal and temporal lobes mediates internal forward models during speech production. *Brain Lang.* **240**, 105266 (2023).
33. Li, S., Zhu, H. & Tian, X. Corollary discharge versus efference copy: distinct neural signals in speech preparation differentially modulate auditory responses. *Cereb. Cortex* **30**, 5806–5820 (2020).
34. Buzsáki, G. & Wang, X.-J. Mechanisms of gamma oscillations. *Annu. Rev. Neurosci.* **35**, 203–225 (2012).
35. Besserve, M., Schölkopf, B., Logothetis, N. K. & Panzeri, S. Causal relationships between frequency bands of extracellular signals in visual cortex revealed by an information theoretic analysis. *J. Comput. Neurosci.* **29**, 547–566 (2010).
36. Crone, N. E., Sinai, A. & Korzeniewska, A. High-frequency gamma oscillations and human brain mapping with electrocorticography. *Prog. Brain Res.* **159**, 275–295 (2006).
37. Miller, R. Theory of the normal waking EEG: from single neurones to waveforms in the alpha, beta and gamma frequency ranges. *Int. J. Psychophysiol.* **64**, 18–23 (2007).
38. Fries, P. Neuronal gamma-band synchronization as a fundamental process in cortical computation. *Annu. Rev. Neurosci.* **32**, 209–224 (2009).
39. Fries, P. Rhythms for cognition: communication through coherence. *Neuron* **88**, 220–235 (2015).
40. Canolty, R. T. et al. Spatiotemporal dynamics of word processing in the human brain. *Front. Neurosci.* **1**, 78 (2007).
41. Crone, N. E., Miglioretti, D. L., Gordon, B. & Lesser, R. P. Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. *Brain J. Neurol.* **121**, 2301–2315 (1998).
42. Edwards, E., Soltani, M., Deouell, L. Y., Berger, M. S. & Knight, R. T. High gamma activity in response to deviant auditory stimuli recorded directly from human cortex. *J. Neurophysiol.* **94**, 4269–4280 (2005).
43. Edwards, E. et al. Comparison of time–frequency responses and the event-related potential to auditory speech stimuli in human cortex. *J. Neurophysiol.* **102**, 377–386 (2009).
44. Towle, V. L. et al. ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. *Brain* **131**, 2013–2027 (2008).
45. Manning, J. R., Jacobs, J., Fried, I. & Kahana, M. J. Broadband shifts in local field potential power spectra are correlated with single-neuron spiking in humans. *J. Neurosci.* **29**, 13613–13620 (2009).
46. Miller, K. J. et al. Human Motor Cortical Activity Is Selectively Phase-Entrained on Underlying Rhythms. *PLOS Comput. Biol*. **8**, 1–21 (2012).
47. Donoghue, T. et al. Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* **23**, 1655–1665 (2020).
48. Ray, S., Hsiao, S. S., Crone, N. E., Franaszczuk, P. J. & Niebur, E. Effect of stimulus intensity on the spike-local field potential relationship in the secondary somatosensory cortex. *J. Neurosci.* **28**, 7334–7343 (2008).
49. Jia, X. & Kohn, A. Gamma rhythms in the brain. *PLoS Biol.* **9**, e1001045 (2011).
50. Ray, S. & Maunsell, J. H. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol.* **9**, e1000610 (2011).
51. Rich, E. L. & Wallis, J. D. Spatiotemporal dynamics of information encoding revealed in orbitofrontal high-gamma. *Nat. Commun.* **8**, 1139 (2017).
52. Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M. & Gagnon, D. A. Lexical access in aphasic and nonaphasic speakers. *Psychol. Rev.* **104**, 801 (1997).
53. Jacquemot, C. & Scott, S. K. What is the relationship between phonological short-term memory and speech processing?. *Trends Cogn. Sci.* **10**, 480–486 (2006).
54. Levelt, W. J., Praamstra, P., Meyer, A. S., Helenius, P. & Salmelin, R. An MEG study of picture naming. *J. Cogn. Neurosci.* **10**, 553–567 (1998).
55. Cai, Q. & Brysbaert, M. SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PloS One* **5**, e10729 (2010).
56. Papademetris, X. et al. BioImage Suite: an integrated medical image analysis suite: an update. *Insight J.* **2006**, 209 (2006).
57. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* **2011**, 1–13 (2011).
58. Shattuck, D. W. & Leahy, R. M. BrainSuite: an automated cortical surface identification tool. *Med. Image Anal.* **6**, 129–142 (2002).
59. Joshi, A. A. et al. A whole brain atlas with sub-parcellation of cortical gyri using resting fMRI. in *Medical Imaging 2017: Image Processing* (eds Styner, M. A. & Angelini, E. D.) vol. 10133 101330O (SPIE, 2017).
60. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* **134**, 9–21 (2004).
61. Morgan, A. M. et al. Decoding words during sentence production with ECoG reveals syntactic role encoding and structure-dependent temporal dynamics. *Commun. Psychol.* **3**, 87 (2025).
62. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B Methodol.* **57**, 289–300 (1995).
63. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
64. Diks, C. & Panchenko, V. A new statistic and practical guidelines for nonparametric Granger causality testing. *J. Econ. Dyn. Control* **30**, 1647–1669 (2006).
65. Lütkepohl, H. *Introduction to Multiple Time Series Analysis* (Springer Science & Business Media, 2013).
66. Sameshima, K. & Baccalá, L. A. asympPDC Package. *GitHub* https://github.com/asymppdc/asympPDC (2022).
67. Lu, Z. & Ku, Y. Neurora: a Python toolbox of representational analysis from multi-modal neural data. *Front. Neuroinform.***14**, 563669 (2020).
68. waws520waws. ttskit: Text To Speech Toolkit (v0.1.2). *GitHub*. https://github.com/waws520waws/ttskit (2022).
69. Altmann, C. F. et al. Temporal dynamics of adaptation to natural sounds in the human auditory cortex. *Cereb. Cortex* **18**, 1350–1360 (2008).
70. Correia, J. M., Jansma, B. M. & Bonte, M. Decoding articulatory features from fMRI responses in dorsal speech regions. *J. Neurosci.* **35**, 15015–15025 (2015).
71. Gong, X. L. et al. Phonemic segmentation of narrative speech in human cerebral cortex. *Nat. Commun.* **14**, 4309 (2023).

72. The MathWorks Inc. *MATLAB Version: 9.9.0 (R2020b)* (The MathWorks Inc., Natick, Massachusetts, 2020).

Acknowledgements
We thank Lu Luo, Hao Zhu, Yuchunzi Wu, and Yunzhe Sun for their assistance in data collection and analysis. This study was supported by the National Natural Science Foundation of China (32071099 and 32271101 to X.T., 32441106 and 32171039 to Q.W.), National Science and Technology Innovation 2030 Major Program (2022ZD0204804 to Q.W.), Program of Introducing Talents of Discipline to Universities Base B16018, NYU Shanghai Boost Fund, Shanghai Frontiers Science Center of Artificial Intelligence and Deep Learning at NYU Shanghai, NYU Shanghai Summer Undergraduate Research Experience Program Fund (to Z.C.), NYU Shanghai Capstone Fund (to Z.C.), NYU Grossman School of Medicine Vilcek Institute Travel Grant (to Z.C.), and NYU Joint Neuroscience Program Travel Grant (to Z.C.).

Author contributions
S.L. and X.T. conceived the study and designed the experiments; S.L., J.W., P.T., Q.W., and G.L. conducted the research; S.L., Z.C., and X.L. performed data analyses. S.L. and Z.C. wrote the initial draft of the manuscript; S.L., Z.C., X.L., Q.W., and X.T. edited and reviewed the final manuscript, and all authors read and approved the manuscript. Q.W. and X.T. acquired funding; Q.W. and X.T. supervised the study.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-025-08847-6.

Correspondence and requests for materials should be addressed to Qian Wang or Xing Tian.

Peer review information *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editor: Jasmine Pan.