



Revealing the spatiotemporal brain dynamics of covert speech compared with overt speech: A simultaneous EEG-fMRI study

Wei Zhang^{a,b,1}, Muyun Jiang^{c,1}, Kok Ann Colin Teo^{a,b,e,h}, Raghavan Bhuvanakantham^{a,b}, LaiGuan Fong^a, Wei Khang Jeremy Sim^{a,e}, Zhiwei Guo^c, Chuan Huat Vince Foo^d, Rong Hui Jonathan Chua^d, Parasuraman Padmanabhan^{a,b}, Victoria Leong^{f,g}, Jia Lu^{a,d,i}, Balázs Gulyás^{a,b,j,**}, Cuntai Guan^{c,*}

^a Cognitive Neuroimaging Centre, Nanyang Technological University, Singapore

^b Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore

^c School of Computer Science and Engineering, Nanyang Technological University, Singapore

^d DSO National Laboratories, Singapore

^e IGP-Neuroscience, Interdisciplinary Graduate Programme, Nanyang Technological University, Singapore

^f Division of Psychology, Nanyang Technological University, Singapore

^g Department of Pediatrics, University of Cambridge, United Kingdom

^h Division of Neurosurgery, National University Health System, Singapore

ⁱ Yong Loo Lin School of Medicine, National University of Singapore, Singapore

^j Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Keywords:

Covert speech
Overt speech
Simultaneous EEG-fMRI
Left putamen
Functional connectivity

ABSTRACT

Covert speech (CS) refers to speaking internally to oneself without producing any sound or movement. CS is involved in multiple cognitive functions and disorders. Reconstructing CS content by brain-computer interface (BCI) is also an emerging technique. However, it is still controversial whether CS is a truncated neural process of overt speech (OS) or involves independent patterns. Here, we performed a word-speaking experiment with simultaneous EEG-fMRI. It involved 32 participants, who generated words both overtly and covertly. By integrating spatial constraints from fMRI into EEG source localization, we precisely estimated the spatiotemporal dynamics of neural activity. During CS, EEG source activity was localized in three regions: the left precentral gyrus, the left supplementary motor area, and the left putamen. Although OS involved more brain regions with stronger activations, CS was characterized by an earlier event-locked activation in the left putamen (peak at 262 ms versus 1170 ms). The left putamen was also identified as the only hub node within the functional connectivity (FC) networks of both OS and CS, while showing weaker FC strength towards speech-related regions in the dominant hemisphere during CS. Path analysis revealed significant multivariate associations, indicating an indirect association between the earlier activation in the left putamen and CS, which was mediated by reduced FC towards speech-related regions. These findings revealed the specific spatiotemporal dynamics of CS, offering insights into CS mechanisms that are potentially relevant for future treatment of self-regulation deficits, speech disorders, and development of BCI speech applications.

Major abbreviations

AAL automated anatomical labeling
AAS average artifact template subtraction
BCG ballistocardiogram

BCI brain-computer interface
BEM boundary element model
BOLD blood oxygenation level-dependent
CFI comparative fit index
CS covert speech

* Corresponding authors at: School of Computer Science and Engineering, Nanyang Technological University, Singapore.

** Corresponding authors at: Director of Cognitive Neuroimaging Centre, Nanyang Technological University, Singapore.

E-mail addresses: balazs.gulyas@ntu.edu.sg (B. Gulyás), ctguan@ntu.edu.sg (C. Guan).

¹ These authors contribute equally to this work.

<https://doi.org/10.1016/j.neuroimage.2024.120629>

Received 5 December 2023; Received in revised form 17 April 2024; Accepted 29 April 2024

Available online 30 April 2024

1053-8119/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

ECG	electrocardiogram
ECoG	electrocorticogram
EEG	electroencephalography
EPI	echo-planar imaging
FC	functional connectivity
FDR	false discovery rate
fMRI	functional magnetic resonance imaging
FWHM	full width at half maximum
GLM	general linear model
HRF	hemodynamic response function
IC	independent component
ICA	independent component analysis
MEG	magnetoencephalography
MNI	Montreal Neurological Institute
MPRAGE	Magnetization Prepared Rapid Gradient Echo
OS	overt speech
PCG	precentral gyrus
PET	positron emission tomography
PSD	power spectral density
RMSEA	root mean square error of approximation
ROI	region of interest
RS	resting state
SMA	supplementary motor area
SPM	Statistical Parametric Mapping
SRMR	standardized root mean square residual
TE	echo time
TLI	Tucker-Lewis Index
TR	repetition time

1. Introduction

Covert speech (CS), also known as inner speech, refers to speaking internally to oneself without producing any sound or movement. CS is believed to be associated with multiple cognitive functions, such as memory, control, planning, reading, writing, and self-awareness (Alderson-Day and Fernyhough, 2015; Lidstone et al., 2010; Marvel and Desmond, 2012). CS also plays an important role in self-regulation, in the situations like impulse and emotional regulation and delayed reinforcement (Barkley, 1997; Day and Smith, 2013; Tullett and Inzlicht, 2010). The ability of CS can be evaluated by self-reports or by the performance during specialized covert language tests, and the deficit of CS underlies speech disorders like stuttering, auditory verbal hallucinations, and aphasia (Brocklehurst and Corley, 2011; Fernyhough et al., 2019; Langland-Hassan et al., 2015), and mental diseases like autism and schizophrenia (Albein-Urios et al., 2021; Barber et al., 2021). Reconstructing CS content by brain-computer interface (BCI) also offers the potential to restore communication functions when hindered by disease or environmental factors (Cooney et al., 2018; Gonzalez-Lopez et al., 2020; Proix et al., 2022).

It is necessary to understand the neural mechanisms of CS. Current theories posit that, from behavioral view, CS can be seen as truncated overt speech (OS) without motor execution and actual sensory feedback (Oppenheim and Dell, 2010). However, the neural mechanisms underlying CS compared with OS remain poorly understood (Cooney et al., 2018).

Neuroimaging techniques, such as functional magnetic resonance imaging (fMRI) and positron emission tomography (PET), have identified shared brain regions between CS and OS (including typical dominant-hemispheric language regions), which suggests that CS and OS exist within the same continuum of neural processes, although the difference in neural activity strengths is debated (Alderson-Day and Fernyhough, 2015; Basho et al., 2007; Perrone-Bertolotti et al., 2014). In contrast, other studies argue that CS might involve independent neural processes based on the preservation of CS ability in aphasia patients (Geva et al., 2011; Stark et al., 2017). Beyond the debates of involved regions, neuroelectrophysiology studies detect distinct brain dynamics

patterns between CS and OS, although in lower spatial precision (Burle et al., 2015). For example, different neural activity strengths were found between CS and OS in EEG (Stephan et al., 2020) and magnetoencephalography (MEG) (Goto et al., 2011; Voigtlaender et al., 2023) during various stages of speech production. Furthermore, different patterns were detected by electrocorticogram (ECoG) (Martin et al., 2014; Pei et al., 2011). In summary, a more inclusive view between CS and OS is that they can engage a shared network in different extents and might involve independent pathways with distinct dynamics.

However, as suggested in a previous review (Martínez-Manrique and Vicente, 2015), more comprehensive neuroimaging data are needed to precisely understand the distinct neural mechanisms between CS and OS. This is because single modalities have limitations in spatial resolution (EEG, MEG), temporal resolution (fMRI) and feasibility (the invasive ECoG) (Kim et al., 1997; Lopes da Silva, 2013; Parvizi and Kastner, 2018). A promising approach forward is multi-modal analysis, using simultaneous EEG-fMRI imaging with the techniques such as fMRI-informed EEG source localization (Huster et al., 2012). Currently, although there are some non-simultaneous EEG-fMRI studies investigating the decoding of CS content by BCI (Simistira Liwicki et al., 2023; Wilson et al., 2023; Yoshimura et al., 2016), the temporal signal-to-noise ratio is significantly lower as compared with the simultaneous EEG-fMRI (Schrooten et al., 2018). To the best of our knowledge, the neural mechanisms of CS have not been directly explored through the technique of simultaneous EEG-fMRI, which can provide precise spatio-temporal insights into the ongoing debate of distinct CS and OS mechanisms, and might help to optimize the timing and targeting of future language-related treatments (Abo et al., 2012; Musso et al., 2022; Stoeckel et al., 2014). In this study, we hypothesized that simultaneous EEG-fMRI can potentially reveal the spatiotemporal dynamics of CS. If CS and OS differ in activations observed in specific regions or specific time window, it will support the involvement of unique CS pathways, and vice versa. To test this hypothesis, we performed a simultaneous fMRI-EEG experiment including both CS and OS tasks. We extracted spatiotemporal features and analyzed their effect pathways using functional connectivity and multivariate path analysis. We expect that these findings might offer insights into the mechanisms of CS and potential neural features for the future treatments of speech disorders and the development of BCI speech applications.

2. Methods

2.1. Participants

We applied the following inclusion criteria to maintain data homogeneity. Only males were included, as numerous studies reported significant sex differences in neural activity during speech (Sato, 2020; Shaywitz et al., 1995; Zaidi, 2010; Hirnstein et al., 2019). Participants aged between 20 and 40 years were selected to ensure homogeneity in the healthy status of brain anatomy and function. Only right-handed native English speakers were included to account for homogeneity in language-related brain features.

The exclusion criteria included: 1. MR contraindications. 2. A lifetime history of psychotic, bipolar, or substance-use disorders. 3. Depressive or anxiety disorders within the past year. 4. Any neurological or significant somatic illnesses. 5. The use of psychotropic medications or other agents known to affect cognitive functions.

As a result, 34 healthy male participants were initially included in this experiment. Two of the participants were excluded because they did not complete the study. Finally, 32 participants completed the experiment, falling within the age range of 24.1 ± 3.9 years.

2.2. Study design

2.2.1. Hierarchical structure of the experiment design

We performed a word-speaking experiment with simultaneous EEG-

fMRI recording. The structure of the experimental protocol is hierarchical (Fig. 1a). The experiment is organized into three main types of units, from largest to smallest: blocks, runs, and trials:

- 1) The experiment comprised 1 block (5.5 min) of resting state (RS), 5 blocks (5 min each) of CS tasks, and 5 blocks (5 min each) of OS tasks. CS and OS blocks alternated with each other during the experiment. Rest interval between blocks is 45 s.
- 2) Each CS/OS block consisted of 20 runs. Rest interval between runs is 3 s.
- 3) Each run consists of 5 trials. Rest interval between trials is 0.5 s.
- 4) A trial is the smallest unit of the experiment, lasting 1.5 seconds, during which a single word is displayed and spoken.

2.2.2. Word speech task implementation

In each run, a word was randomly selected from a set of five words and repeated across 5 trials within that run.

During the trials within OS blocks, participants were directed to speak each word loudly and clearly, while maintaining the natural tone of their daily speech. During the trials within CS blocks, participants were instructed to imagine the muscle movements of speaking the word without any physical movement or producing any sound.

The five-word set was designed based on the following criteria, as

suggested in a previous review (Cooney et al., 2018). 1. Appropriate length with multiple syllables. 2. Distinct pronunciations between words. 3. Frequent usage in BCI scenarios, mirroring real-world BCI applications. The chosen words were "Go there", "Distract Target", "Follow me", "Explore Here", and "Terminate". In total, each word was repeated 100 times in random order separately during CS and OS.

2.2.3. Experimental settings

For the experimental settings, participants were situated inside the MRI scanner with simultaneous EEG recording devices. A mirror was mounted at their forehead level within the scanner, allowing them to view the screen in a straight-ahead gaze. The screen, which was positioned behind the scanner bore, displayed speech task cues and experiment instructions using PsychoPy (Peirce et al., 2019). The screen instructions of the whole experiment are detailed in Supplementary Material 1 with both figure and video demos.

The study was in accordance with the Declaration of Helsinki and was conducted in Cognitive Neuroimaging Centre (CoNiC) in Nanyang Technological University (NTU) in Singapore. All the participants provided written informed consent. Ethics committee approval was obtained from the Institutional Review Board (IRB) of NTU, with the approval number: IRB-2022-040.

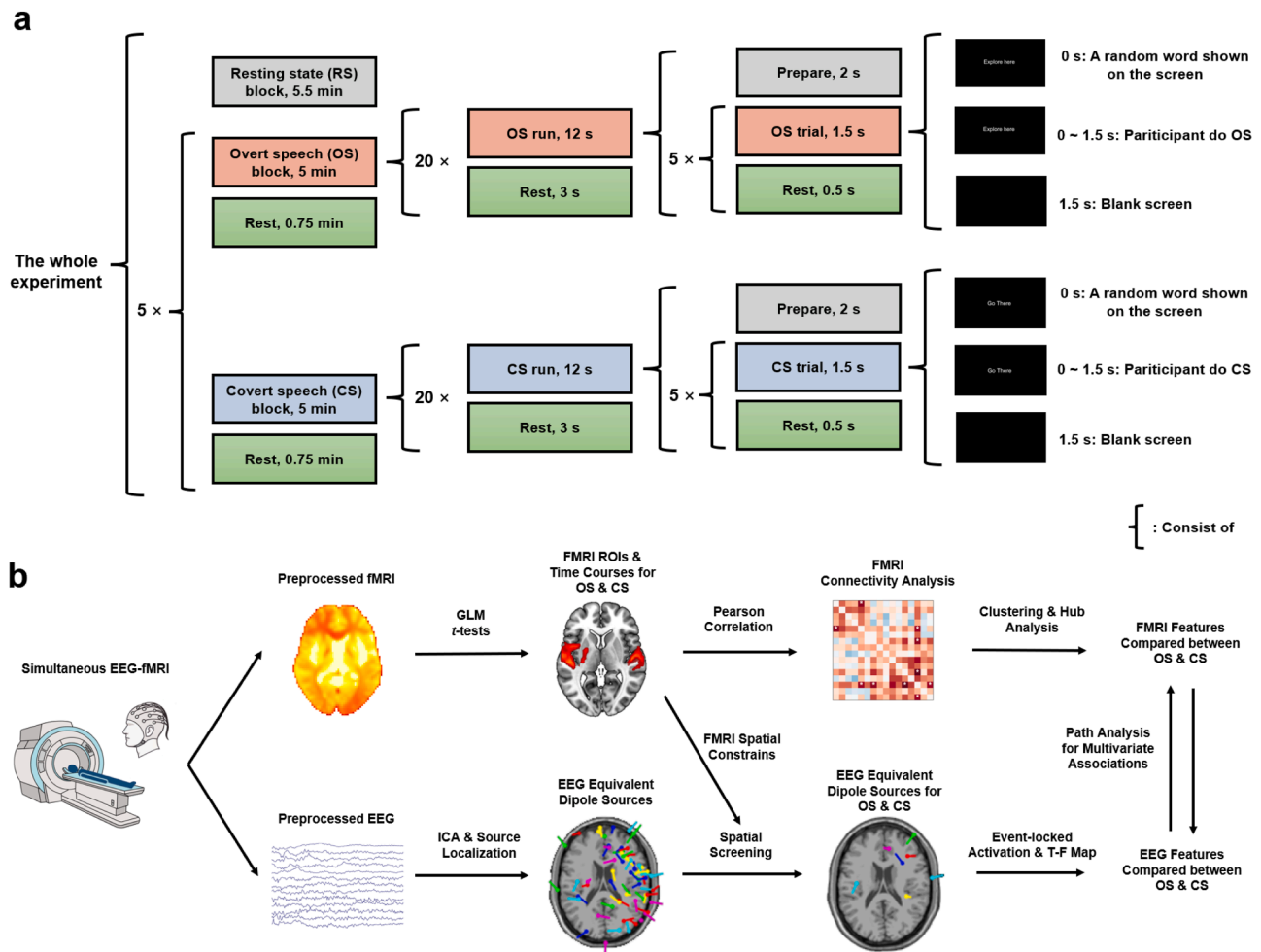


Fig. 1. Flow chart of the study. (a) The structure of the experiment in hierarchical order. The experiment consists of 1 resting state (RS) blocks, 5 covert speech (CS) blocks, 5 overt speech (OS) blocks. CS and OS blocks alternate with each other during the experiment. Each block consists of 20 runs, and each run consists of 5 trials. A trial is the smallest unit of the experiment, lasting 1.5 seconds, during which a single word is displayed and spoken. Rest intervals are set between blocks, runs, and trials. (b) The analyzing pipeline of the simultaneous EEG-fMRI data. Except for the basic single-modality analyses, we performed the method of fMRI-informed EEG source localization for multi-modal analysis, which is constraining and screening the EEG dipole sources using the regions of interest (ROIs) extracted from fMRI analysis. At the final step, we performed the path analysis to statistically examine the hypothesized multivariate associations between various features.

2.3. Data acquisitions

2.3.1. EEG acquisitions

EEG was recorded continuously throughout the experiment alongside fMRI acquisition, using the same EEG hardware settings as the previous study (Andreou et al., 2017). We used MR-compatible devices of BrainVision Recorder (Version 1.10, Brain Products, Munich, Germany), AC-amplifiers specifically designed for EEG in MRI scanner (BrainAmp MRplus; Brain Products), and an electrode cap with 64 ring-type electrodes (BrainCapMR 64, Brain Products). The cap was equipped with 62 EEG electrodes based on a modified 10/10 system, with the reference electrode of FCz, the ground electrode of AFz, and an electrocardiogram (ECG) electrode on the participants' back. The cable connecting the electrode wires to the amplifiers was positioned using sand weights atop foam pads to decrease artifacts from the scanner's oscillations. The impedance between the electrode and skin remained under 10 k Ω by manual inspection. Data were initially sampled at a high frequency of 5000 Hz for the MRI artifact removal during preprocessing. The amplitude resolution is 0.5 μ V. Event and fMRI marker information were automatically embedded in EEG data as annotations.

2.3.2. MRI acquisitions

MRI scans were continuous throughout the experiment, performed using a 3T Siemens Prisma 750 scanner. fMRI was acquired by measuring blood oxygenation level-dependent (BOLD) signals with an echo-planar imaging (EPI) sequence: Repetition time (TR) = 1500 ms, echo time (TE) = 30 ms, flip angle = 40°, field of view (FOV) = 24.0 \times 24.0 cm, 26 slices with slice thickness = 3.5 mm, and matrix size = 96 \times 96. T1-MRI was acquired for the spatial coregistration of fMRI, using the Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence: TR = 2200 ms, TE = 2.28 ms, inverse time (TI) = 994 ms, flip angle = 8°, FOV = 24.0 \times 25.6 cm, 208 slices with slice thickness = 1 mm, and matrix size = 240 \times 256.

2.4. Data preprocessing

2.4.1. Preprocessing of EEG

We summarized the whole processes and data format of EEG preprocessing in Table 1.

First, the fMRI-induced gradient artifacts and pulse artifacts were corrected in the raw EEG data using an average artifact template subtraction (AAS) algorithm. Ballistocardiogram (BCG) artifacts were corrected based on the ECG channel. The EEG data were average re-referenced and further filtered between 0.5 - 70 Hz. The preprocessing is based on BrainVision Analyzer 2.2 (BrainVision Analyzer, 2021).

Next, EEG data were resampled to 250 Hz, and the independent component analysis (ICA) was performed using EEGLAB (Delorme & Makeig, 2004). The ICLABEL (Pion-Tonachini et al., 2019) was employed to automatically clear independent components (IC) associated with non-brain artifacts (especially speech-motor artifacts). Then, the EEG data were segmented into repeated overt and covert trials, which was followed by excluding bad trials characterized as containing EEG with abnormal signal gradient (> 50 μ V / ms) using BrainVision Analyzer 2.2.

After the above segmentation and exclusion of trials with bad EEG quality, in total 67.79% (10,847 out of 16,000) of OS trials and 77.57% (12,411 out of 16,000) of CS trials were retained, and the trials retained for each participant were shown in Supplementary Material 2.

2.4.2. Preprocessing of fMRI

Task-fMRI images were processed based on Statistical Parametric Mapping 12 (SPM12) (Ashburner et al., 2014), following the procedure of a previous simultaneous EEG-fMRI study (Andreou et al., 2017). It included the processes of removing the first 10 time points, slice timing, realignment, affine registration, spatial normalization to Montreal Neurological Institute (MNI) space based on individual T1-MRI, and

Table 1

EEG data processing pipeline, including detail steps from preprocessing to fMRI-informed source localization and statistical analysis.

Major steps	EEG processing procedures	Tools	EEG data format	
Preprocessing	Raw EEG data	None	Channels \times Experiment time length	
	Clear fMRI & BCG artifacts	BrainVision Analyzer	Channels \times Experiment time length	
	Re-reference, filter, resample	BrainVision Analyzer	Channels \times Experiment time length	
	ICA for detecting artifact	EEGLAB	ICs \times Experiment time length	
	Clear artifact ICs	EEGLAB & ICLABEL	Channels \times Experiment time length	
	Segment into event-locked trials	BrainVision Analyzer	Channels \times Trials \times Trial time length	
	Exclude bad trials	BrainVision Analyzer	Channels \times Valid trials \times Trial time length	
	Concatenate valid trials	EEGLAB	Channels \times Experiment time length	
	Source localization	ICA	EEGLAB	ICs \times Experiment time length
		fMRI-informed source localization	DIPFIT	ICs in fMRI-ROIs \times Experiment time length
Segment into event-locked trials		EEGLAB	ICs in fMRI-ROIs \times Valid trials \times Trial time length	
Statistical analysis in trials		Matlab	ICs in fMRI-ROIs \times Valid trials \times Trial time length	

Abbreviations: BCG = Ballistocardiogram, IC = Independent component, ICA = Independent component analysis

smoothing with a width at half maximum (FWHM) = 4 mm Gaussian kernel. No nuisance covariates regression or linear trend was removed because they were usually used in resting-state fMRI and might result in the removal of the task effects.

2.5. Data analysis

2.5.1. Identifying regions of interest using fMRI with general linear model and T-test

General linear model (GLM) was performed for estimating the effect of different conditions on each series of fMRI images. In GLM, for each voxel of fMRI:

$$S_t = \beta_0 + \beta_1 R_t + \beta_2 O_t + \beta_3 C_t + \varepsilon_t$$

Where t refers to the time points in a fMRI series, S_t refers to the BOLD signal at t , β_0 refers to the intercept of the signal, R_t refers to the label of RS at t , O_t refers to the label of OS at t , C_t refers to the label of CS at t , β_{1-3} refer to the coefficients of different conditions, and ε_t refers to the residuals. The condition labels were adjusted using the canonical Hemodynamic Response Function (HRF) to fit the BOLD response.

We performed voxel-wise T-tests to identify areas with significant BOLD changes during CS and OS compared with RS. A second-level group analysis summarized these findings across all participants, including age as a covariate to account for its effect on BOLD signal (Richter and Richter, 2003). Voxels with $P < 0.05$ (false discovery rate (FDR) corrected for multiple comparisons) were considered significant and used to define the regions of interest (ROIs), with the labeling from Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). Any regions outside the AAL brain mask or smaller than 500 mm³ were excluded to minimize false positives.

2.5.2. fMRI-informed EEG source localization analysis

The data analysis pipeline for simultaneous EEG-fMRI is shown in Fig. 1b. We performed fMRI-informed EEG source localization (Huster et al., 2012) to estimate EEG dipole sources guided by the ROI

constraints from fMRI. It is implemented by DIPFIT (<https://github.com/scn/dipfit>) within EEGLAB, based on FieldTrip (Oostenveld et al., 2011). The EEG pipeline of the source localization process is also included in Table 1.

First, we utilized the standard boundary element model (BEM) template (Oostendorp and Van Oosterom, 1989; Oostenveld et al., 2001) in the MNI space, which was spatially coregistered with the individual T1-MRI image to model the head anatomy. The electrode system was also aligned with the MNI template.

Second, ICA was performed on the concatenated EEG data across all trials for each participant separately on the covert and overt conditions, resulting in 63 ICs per participant. The optimal dipole locations and orientations for each EEG IC was estimated based on the best fit of its equivalent dipole's projection on the scalp electrodes, while single dipole but not symmetric double dipoles was used to better fit the lateralization of speech (Price, 2012). This IC-based source localization approach ensured that the dipole fitting results were based on the full experimental data rather than individual trials, thus improving the signal-to-noise ratio and the reliability of the results. The IC activations were then Z-scored across the entire experiment and baseline-corrected using the mean value during the 0.25 s period before each trial onset. Finally, to identify the EEG dipole sources that were relevant to speech, we used the ROIs from fMRI GLM analysis as the spatial constraints to screen the speech-relevant EEG dipole sources that were contained within, following the implementations of the previous studies (Bak et al., 2011; Meyer et al., 2012; Ullsperger & Von Cramon, 2001).

After identifying EEG sources by fMRI-ROIs, we compared the following EEG features between CS and OS:

To identify significant event-locked EEG activation, we performed Z-tests among all valid trials from all participants, separately in six discrete time bins. These time bins were 0.25 s long, and were distributed within the task period (0 - 1.5 s after the stimulus). These Z-tests aimed to identify time bins where the amplitude was greater than zero with a 95 % confidence level, and correcting *P*-values by FDR. Z-test was more suitable for the large sample size of repeated EEG trials (Abebe, 2019). Both bootstrapping and permutation tests, as non-parametric methods, were performed to re-verify the Z-test results for CS and OS data. Event locked time-frequency analysis was described in Supplementary Material 4.

Spectral analysis of each dipole source was performed by the periodogram power spectral density (PSD) estimate from Matlab. We used 1000 sampling points that were logarithmically spaced between 0.5- 30 Hz. The upper frequency limit was set as 30 Hz to avoid muscular artifacts of OS (Muthukumaraswamy, 2013).

2.5.3. fMRI network analysis

We also conducted functional connectivity (FC) analyses to investigate the network-level differences between CS and OS, complementing the regional findings from fMRI-informed EEG source localization. FC was calculated using Pearson's correlation coefficient between the average BOLD signals within the ROIs. The ROI list was defined by the group-wise T-tests on GLM. We calculated FC within three periods of the tasks: OS, CS, and RS. We performed T-tests on the FC values of "OS - RS", "CS - RS", and "OS - CS" across all participants. Permutation tests were also performed to detect significant FC pairs as a non-parametric re-verification for the T-tests.

We calculated the FC degree of each ROI as a quantifiable measure of the connection centrality of a specific region within the entire network. It is computed by summing the absolute FC values between the chosen ROI and all other ROIs within the network. Hub nodes were defined as the nodes with the Z-scored degree > 2.5 (Power et al., 2013).

To identify the community structures within the FC network that group together regions with stronger FC, we utilized the unsupervised method based on modularity optimization (Blondel et al., 2008). This method was implemented using the Community Detection Toolbox (Community Detection Toolbox - File Exchange - MATLAB Central, n.d.) in

MATLAB. The algorithm was applied to achieve optimal clustering of all regions, with the objective of maximizing the modularity metric across the FC network. The FC strength towards each cluster was calculated by averaging across its regions.

2.5.4. Path analysis

Path analysis (PA) is a statistical method based on structural equation modeling that examines hypothesized multivariate associations between features, potentially suggesting causal relationships (Duffy et al., 1981). PA is implemented using maximum likelihood estimation using Lavaan (Version 0.6–16) (Rosseel, 2012). All EEG and fMRI features were adjusted for age covariates before entering the PA model. We only included significant pathways with the absolute standardized regression weight (β) > 0.1 and *P* < 0.05. This exclusion is necessary to consider both effect size and statistical significance (Cohen, 2013), and control the complexity of PA model to avoid overfitting.

The goodness of fit of PA model is measured by commonly used metrics: Confirmatory fit index (CFI), Tucker-Lewis index (TLI), standardized root mean square residual (SRMR), and root mean square error of approximation (RMSEA) with its 90 % CI, *P*-value for the test of a close fit (the null hypothesis is RMSEA < 0.05). The cutoff of these measurements for a good fit is based on the recommendation (Schermelleh-Engel et al., 2003).

Furthermore, we performed the non-parametric permutation test to verify that the PA model significantly fit the real data, by randomly shuffling the features and evaluating the fit after shuffling.

3. Results

3.1. Identifying regions of interest for covert/overt speech based on fMRI

To identify regions of interest (ROIs) where neural activity strength changed during OS or CS relative to RS as spatial prior for further EEG analyses, we performed voxel-wise T-tests on fMRI images, separately for two contrasts (Fig. 2): "Overt Speech versus Resting State" ("OS - RS") and "Covert Speech versus Resting State" ("CS - RS"). For both contrasts, only voxels with increased BOLD signals passed the T-tests (*P* < 0.05, FDR corrected). Detailed names, coordinates, and related statistical data of these ROIs are extensively listed in Table 2. For OS, multiple areas including primary motor and sensory regions were significantly activated. In contrast, for CS, significant activations were observed in the bilateral supplementary motor areas (SMA). Additionally, ANOVA results were provided in Supplementary Material 3, which verified significant ROIs detected by T-tests, as a complementation that showing difference across all three conditions. Moreover, by averaging the BOLD signals within the ROIs identified in Fig. 2a, we presented the time course of these signals for both CS and OS in Fig. 2b. The BOLD signal fluctuations observed across 20 runs within the OS block, exhibiting clear periodic patterns. In contrast, the average BOLD signal in the CS block did not demonstrate similar periodic fluctuations as observed in OS. After normalization by the rest blocks, the BOLD signal strength for both CS and OS remained above zero, indicating relatively stronger activation compared with the nearest rest intervals.

3.2. Difference in fMRI-informed EEG sources between covert and overt speech

To further analyze the event-locked EEG features in fMRI-defined ROIs, we performed fMRI-informed EEG sources localization and compared various EEG features between CS and OS in these ROIs. CS and OS can partially involve common brain regions as addressed in the introduction section. Therefore, we used all the ROIs for both CS and OS (Table 2) as the constraining regions for EEG source localization. Given the study's focus on CS, only three EEG dipole sources with significant event-locked activations during CS were shown (Fig. 3), and other insignificant dipole sources for CS were listed in Supplementary

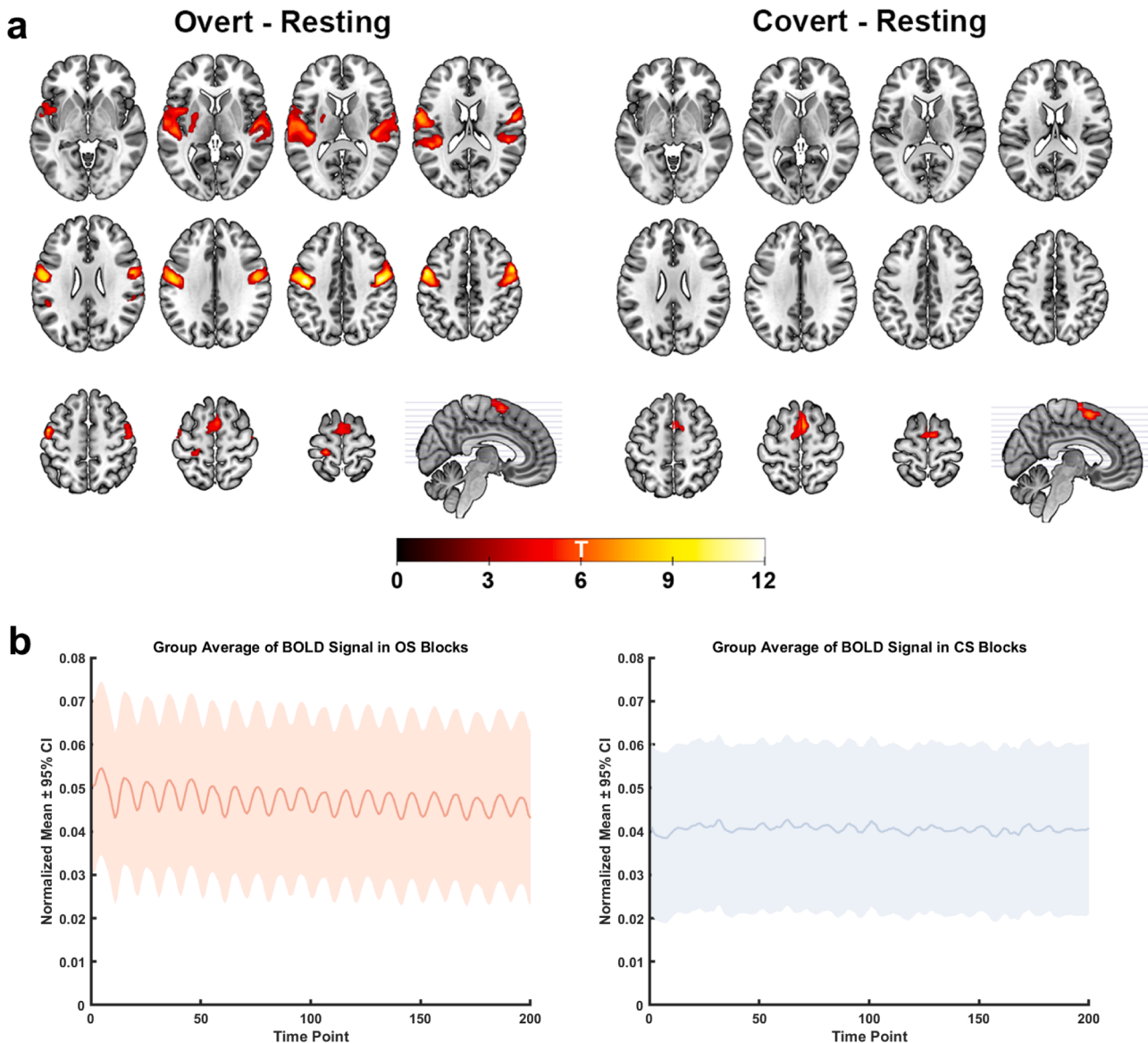


Fig. 2. Neural activity differences between speech tasks and the resting state. (a) BOLD signal difference identified by T-tests on fMRI images. The difference maps are separately as "Overt Speech versus Resting State" ("OS - RS" on the left panel) and "Covert Speech versus Resting State" ("CS - RS" on the right panel). We used the threshold of $P < 0.05$, corrected by false discovery ratio (FDR). No regions with a negative T-value were significant in the T-tests. Significant regions are also listed in Table 2. (b) The average and 95 % confidence interval of BOLD signal strength during OS and CS blocks, calculated by averaging across all repetitions within these blocks and across all participants. The ROIs are defined by the combined set identified in Fig. 2a. For each participant, BOLD signal values are normalized using the values from the rest blocks following each speech block.

Material 4. For the sample size in the event-locked EEG activity analysis, 15 participants with 1288 OS trials and 3282 CS trials had significant activation at the left precentral gyrus (PCG) (Fig. 3a); 11 participants with 3064 OS trials and 1929 CS trials had significant activation at the left SMA (Fig. 3b); 10 participants with 1226 OS trials and 2907 CS trials had significant activation at the left putamen (Fig. 3c), demonstrating considerable trial number for statistical analysis between CS and OS.

Here, we presented the event-locked activation, the PSD, and the topographical map (topo-plot) as the key features of the EEG dipole sources for CS in Fig. 3 and compared them with OS.

Significant event-locked activations were identified using Z-tests in six time bin within the trials segmented by task events (middle column in Fig. 3). The plot of a trial ranged from -0.25 s (pre-stimulus baseline) to 0 s (word stimulus to start speech) to 1.5 s (end of the speech period). The left putamen showed earlier event-locked activations for CS compared with OS (Fig. 3c). Specifically, CS activations were significant

in the 0.25 - 0.5 s time bin (Z -value = 1.896, $P = 0.0382$, corrected by FDR), reaching the peak at 262 ms. In contrast, OS showed significant activation in the subsequent time bin of 0.5 - 1.5 s (Z -value = 3.419, $P < 0.0001$, corrected by FDR), reaching the peak at 1170 ms. In the left PCG and SMA, event-locked activations during CS were consistently weaker than those during OS, with no unique activations specific to CS observed across the time bins. We also performed time-frequency analysis as a supplement to the event-locked activation (see Supplementary Material 4), with significant difference between CS and OS in event-related spectral perturbation (ERSP) and inter-trial coherence (ITC) in the three key regions of Fig. 3.

Additionally, bootstrapping tests were performed to re-verify the significant activations from Z-tests (Supplementary Material 5). Permutation tests were also performed to assess the group difference of amplitudes between CS and OS (Supplementary Material 5). Both non-parametric tests corroborated the Z-test findings.

Table 2

Significant regions of interest (ROIs) in speech tasks relative to resting state, identified by T-tests on fMRI images. This table lists the ROIs in Fig. 2a that exhibited significant changes in the blood oxygen level-dependent (BOLD) signal of fMRI images during both overt and covert speech tasks, compared with resting state. We used the threshold of $P < 0.05$, corrected by false discovery ratio (FDR). Each ROI is defined based on the Automated Anatomical Labelling (AAL) atlas (Tzourio-Mazoyer et al., 2002) with the coordinates in the Montreal Neurological Institute (MNI) space.

Tasks	Region of Interest (ROI)	Cluster Center Voxel			Peak T-value	Cluster Size (mm ³)	
		x	y	z			
Overt Speech	Left Precentral Gyrus (L_PreCG)	-51	-3	45	11.406	6264	
	Right Precentral Gyrus (R_PreCG)	45	-12	39	11.667	7155	
	Right Middle Frontal Gyrus (R_MFG)	51	3	51	7.061	810	
	Left Inferior Frontal Gyrus (L_IFGoperc)	-48	3	6	5.489	567	
	Left Rolandic Operculum (L_ROL)	-42	-24	12	7.008	4239	
	Right Rolandic Operculum (R_ROL)	54	-6	18	6.954	2970	
	Left Supplementary Motor Area (L_SMA)	0	3	66	5.935	2187	
	Right Supplementary Motor Area (R_SMA)	3	6	63	6.267	1782	
	Left Insula (L_INS)	-39	3	3	6.185	1350	
	Left Postcentral Gyrus (L_PoCG)	-42	-12	39	11.717	7533	
	Right Postcentral Gyrus (R_PoCG)	48	-12	36	10.690	3294	
	Left Putamen (L_PUT)	-27	-15	3	5.387	999	
	Left Heschl Gyrus (L_HES)	-36	-27	12	7.626	1323	
	Right Heschl Gyrus (R_HES)	48	-15	6	7.776	1755	
	Left Superior Temporal Gyrus (L_STG)	-48	-18	6	7.751	7587	
	Right Superior Temporal Gyrus (R_STG)	48	-33	18	6.768	4698	
	Covert Speech	Left Supplementary Motor Area (L_SMA)	0	12	60	6.043	2484
		Right Supplementary Motor Area (R_SMA)	6	6	60	7.347	2430

For the PSD analysis (right column in Fig. 3), CS had weaker PSD than OS in a wide range of frequency (0.5 - 30 Hz) in all the three regions in Fig. 3 and in most of the insignificant regions (Supplementary Material 4). We also re-verified the power difference by bar plots and T-tests in each power band separately (Supplementary Material 6).

The topo-plots of EEG dipole sources between OS and CS were highly different in the left putamen (left column in Fig. 3c), showing a strong frontal-occipital contrast in OS and a moderate bilateral-temporal activation in CS. In contrast, the topo-plots for the left PCG and SMA showed more subtle differences between CS and OS.

3.3. Difference in fMRI functional connectivity between overt and covert speech

To explore the differential functional network patterns underlying CS and OS, we calculated functional connectivity (FC) between the 16 ROIs (Table 2) derived from fMRI. T-tests were performed to compare the differences in FC values between OS/CS and RS, as well as between OS and CS (Fig. 4a). Significant FCs were identified using a threshold of Bonferroni-corrected $P < 0.05$. Permutation tests for FC were performed between speech tasks and RS, which re-verified the T-test results in Fig. 4a (Supplementary Material 7). These FC analyses aimed to test our hypothesis that regions with distinct roles in CS should exhibit different

connectivity patterns compared with OS, providing further evidence for their specific involvement in CS networks.

Within the two FC networks, nodes with Z-scored degrees > 2.5 are defined as hub nodes (Power et al., 2013). The left putamen emerged as a prominent hub node in the FC networks of both CS and OS, with Z-scored degrees of 2.589 during OS and 2.755 during CS (lower panel in Fig. 4a). Notably, other regions all showed Z-scored degrees < 1 .

Through unsupervised algorithm of community clustering, ROIs were divided into two distinct functional clusters with strong intra-connectivity. The groupings of the two clusters were the same after performing the algorithm separately in CS and OS networks. As shown in Fig. 4b, the left cluster (L_CLU) includes left-hemispheric regions of precentral, postcentral, Rolandic operculum, inferior frontal, Heschl and superior temporal gyrus. The right cluster (R_CLU) includes other regions mainly in the right hemisphere, along with the left SMA, insula and putamen. Notably, the left putamen bridged both clusters by significant FCs, highlighting its role as the hub node (a cluster view in Fig. 4c).

After identifying the central role of the left putamen and the two separate functional clusters, we compared all the potential contrasts of FC strength between OS and CS. In Fig. 4d, when using two-sample T-tests and considering multiple comparisons by FDR, only FCs between the left putamen and the L_CLU passed the T-test with $t(31) = 2.868$ and corrected P -value = 0.0365. FCs between other regions had insignificant corrected P -values of 0.0595, 0.0893, 0.1526, and 0.0893.

3.4. Exploring multivariate associations between EEG dynamics, fMRI functional connectivity and covert speech

To explore the associations between the key features and CS, we performed path analysis (PA) on the features of EEG dipole sources and FC measures. We built the hypothesized PA model in a data-driven manner and used the backward steps to explore the valid pathways. The detailed steps for building PA model were shown in Supplementary Material 8.

For the valid pathways in the PA model (Fig. 5), first, we identified direct positive effect from late EEG activation in the left putamen towards the OS (standardized regression weight (β) = -0.248, $P < 0.001$), but insignificant effect from early EEG activation in the left putamen towards the CS label ($|\beta| < 0.1$). Second, the early EEG activation in the left putamen was negatively associated with its FC towards bilateral functional clusters ($\beta = -0.103$, $P < 0.01$; $\beta = -0.158$, $P < 0.001$), while late EEG activation had no significant effect on the FC ($|\beta| < 0.1$). Third, we identified the indirect positive effect from early EEG activation on CS label, through the mediation of the FC strength between the left putamen and the left cluster ($\beta = -0.441$, $P < 0.001$), but no indirect mediation effect through the FC strength between the left putamen and the right cluster ($|\beta| < 0.1$).

The resulting PA model fit the data well, based on the suggested goodness-of-fit cutoff (Schermelleh-Engel et al., 2003). Comparative Fit Index (CFI) = 0.988, Tucker-Lewis Index (TLI) = 0.970, Root Mean Square Error of Approximation (RMSEA) = 0.050 (90 % confidence interval = 0.033 - 0.068, with $P = 0.466$ for the close fit null hypothesis), and Standardized Root Mean Square Residual (SRMR) = 0.024. Permutation tests on the PA model were performed 1000 times by randomly shuffling the original data and estimating in the same model. In 1000 permutation iterations, 32 shuffled data yielded lower chi-square values, while 968 yielded higher chi-square values. Therefore, the significance that the model fits real data is $P = 32/1000 = 0.032$. A visualization of the fit metrics across permutations is in Supplementary Material 9.

For the details of the PA model, we provided the raw model including all possible pathways in Supplementary Material 8. The details of regression weight and covariance of the proposed PA model with only significant pathways (Fig. 5) were also listed in Supplementary Material 8.

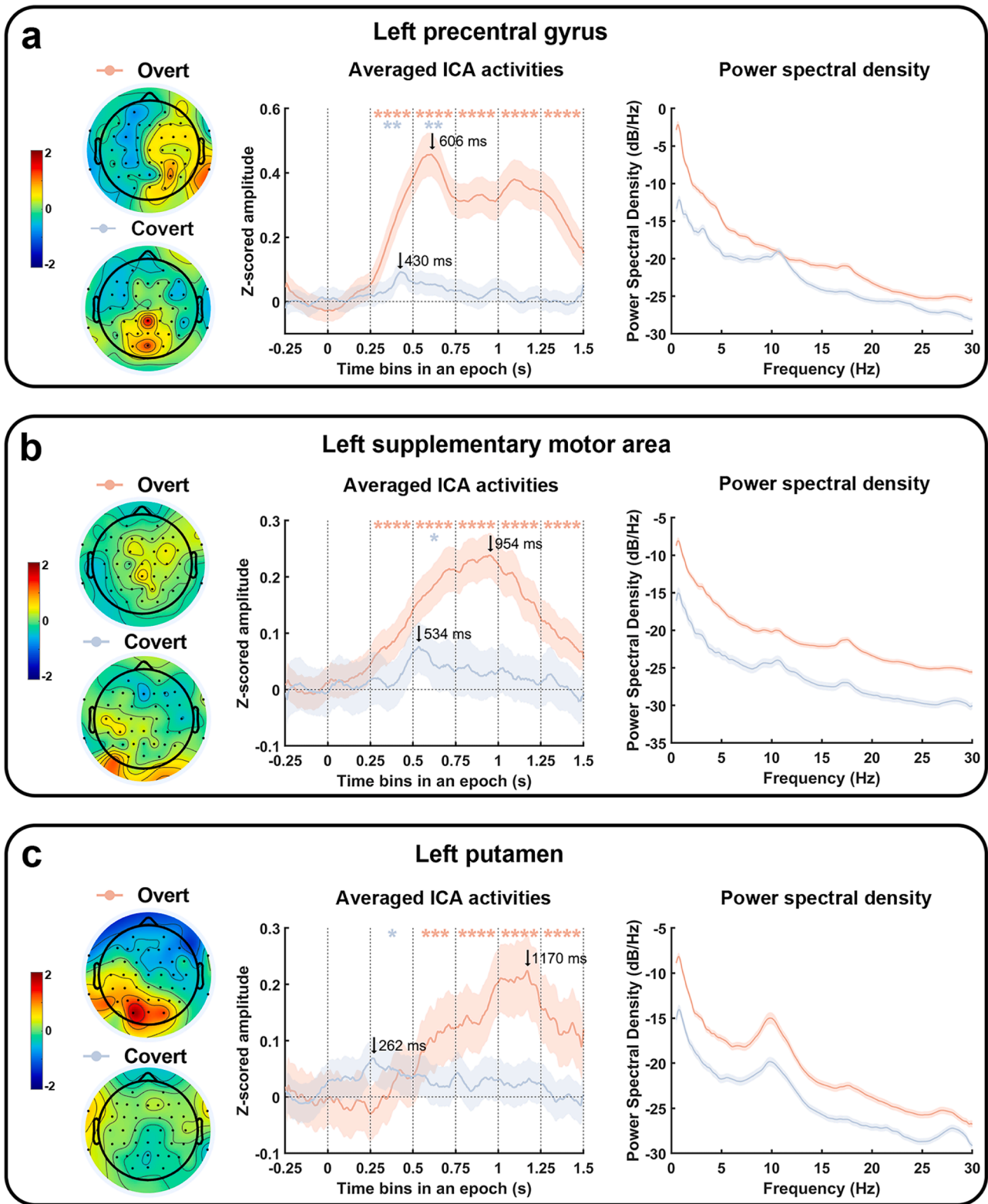


Fig. 3. Differences in EEG features between covert and overt speech tasks, including three regions of fMRI-informed EEG dipole sources that significantly activated during covert speech. Left column: The averaged scalp plot of EEG dipole sources, with outliers beyond triple standard deviations removed; Middle column: Z-scored amplitude of the EEG dipole sources, normalized by the pre-stimulus baseline period (-0.25 to 0 s); Right column: Power spectral density of the EEG dipole sources. Blue and red represent covert and overt speech, respectively. Blue and red asterisks mark significant activations for both speech types (*: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$, ****: $P < 0.0001$, corrected by false discovery ratio (FDR)). Line plots were smoothed with 20 adjacent points. Shadow indicates the 95 % confidence intervals by standard error of mean (SEM). (a) Features of EEG sources in the left precentral gyrus. (b) Features of EEG sources in the left supplementary motor area. (c) Features of EEG sources in the left putamen.

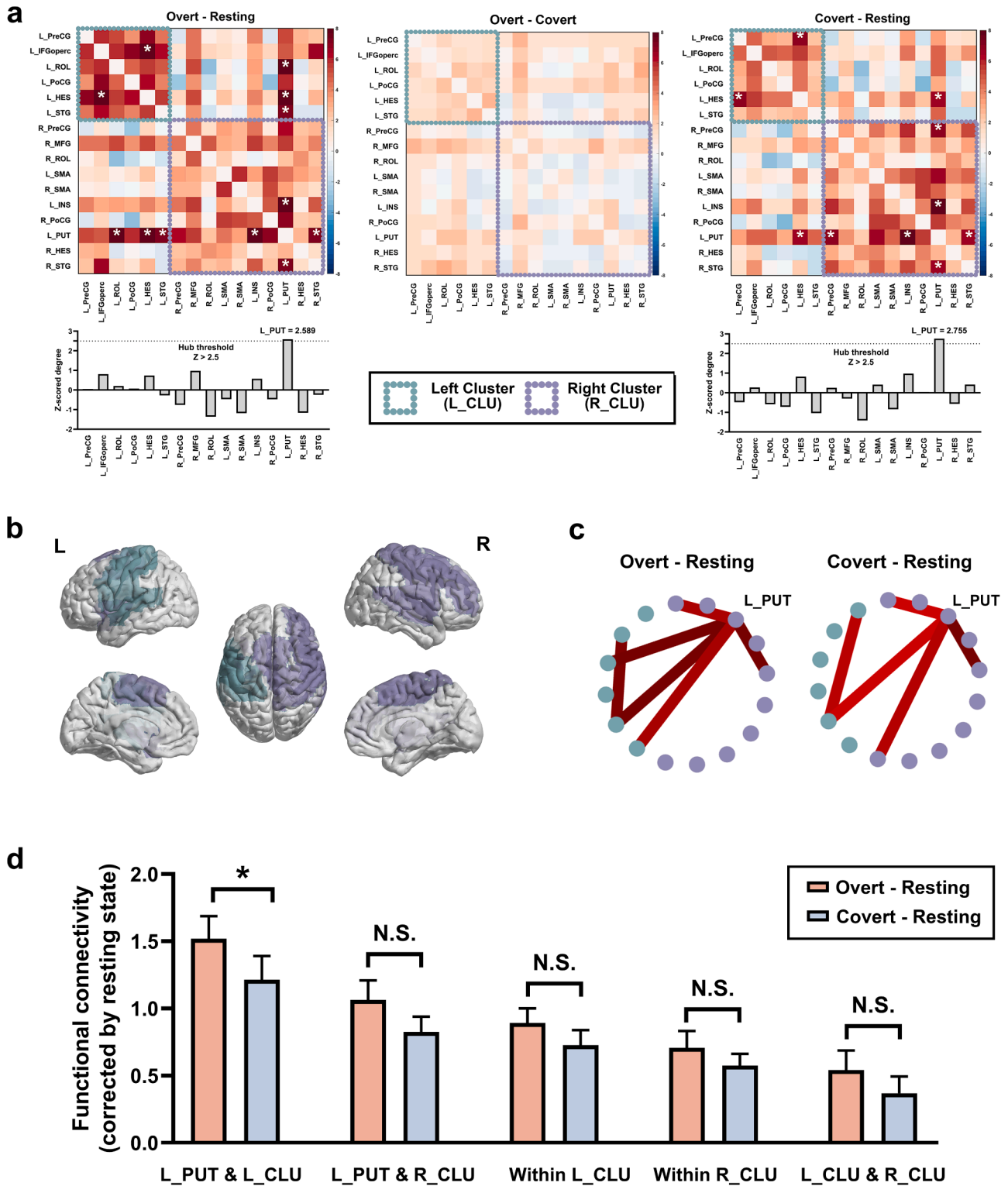


Fig. 4. Differences in fMRI functional connectivity (FC) between overt speech (OS) and covert speech (CS) tasks. Region abbreviations are detailed in Table 2. (a) Upper panel: Adjacency matrices display the FC networks for OS and CS, which are represented by the T-values compared with resting state (RS), as well as the comparison between OS and CS. White asterisks highlight region pairs that have significantly greater FC than RS under T-tests (Bonferroni-corrected $P < 0.05$). Lower panel: Bar plots showing the degree of each node within the adjacency matrices, for OS and CS separately. The left putamen showed as the only hub node in both FC networks. (b) Spatial representation of two identified functional clusters in Fig. 4a, which distributed bilaterally in the brain. (c) Cluster views of network structures highlight the left putamen as a hub node in both OS and CS networks, which also bridged the two clusters by significant FCs. Edge colors are under the same color bar as in Fig. 4a. (d) T-tests showed that FC strengths between the left putamen (L_PUT) and left cluster (L_CLUSTER) was significantly reduced in CS compared with OS ($t(31) = 2.868$ and corrected P -value = 0.0365). Other comparisons were not significant, including between L_PUT and right cluster (R_CLUSTER), within L_CLUSTER, within R_CLUSTER, and between L_CLUSTER and R_CLUSTER. Error bars represent standard error of mean (SEM).

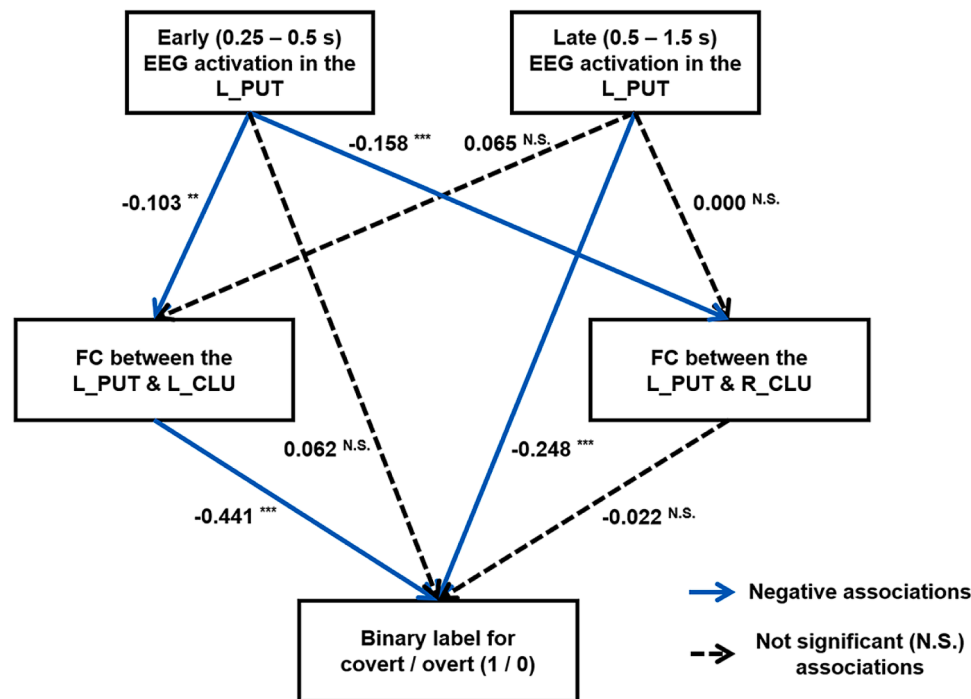


Fig. 5. Path analysis showing the significant associations from EEG dynamics and fMRI functional connectivity (FC) in the left putamen towards the binary classification of covert speech (CS) and overt speech (OS). The model fit the data well based on the suggested cutoff. This model hypothesized and fit the associations among three types of features: 1. Early and late EEG dipole source activations in the left putamen (L_PUT) identified in Fig. 3c; 2. FC between the L_PUT with left (L_CLU) and right clusters (R_CLU) identified in Fig. 4; 3. The binary classification labels of speech trials as covert (marked as 1) or overt (marked as 0). Significant associations (solid blue arrows in the figure) suggest that late EEG activations in the L_PUT have direct effect towards OS. The early EEG activations leads to CS by the mediation effect of reduced FC between the L_PUT and L_CLU. Significance indicators: *** denotes $P < 0.001$, ** denotes $P < 0.01$, and N.S. denotes associations with absolute standardized regression weights below 0.1 or P -values greater than 0.05.

4. Discussion

Our study aims to reveal the spatiotemporal brain dynamics underlying CS compared with OS. We performed a simultaneous EEG-fMRI study on 32 participants, where they repeatedly generated words both overtly and covertly. Based on the multi-modal analysis, we compared the spatiotemporal features between CS and OS, and investigated their effect pathways through functional connectivity and multivariate path analysis. Due to the limited prior knowledge about the neural mechanisms of CS, our statistical analyses were data-driven and exploratory among the multimodal features, inspired by previous studies (Dehghani et al., 2023; Golkowski et al., 2017).

The major findings of this study can be summarized in four aspects: 1. We identified three EEG dipole sources with significant event-locked activations during CS, located in the left PCG, left SMA and left putamen. 2. Although OS involved more brain regions with stronger activations, CS was characterized as the earlier event-locked EEG activation in the left putamen. 3. fMRI FC analysis identified the left putamen as the only hub node of the FC networks during both CS and OS, while showing weaker FC strength towards speech-related regions in the dominant hemisphere during CS. 4. Path analysis focusing on the left putamen verified the indirect effect from early EEG activation towards CS, through the mediation effect of reduced FC towards speech-related regions. Details of results are summarized as table in Supplementary Material 10.

In summary, we propose the potential CS mechanism that the earlier activation of the left putamen might transform OS into CS by inhibiting the FC towards the speech-related network in the dominant hemisphere.

4.1. Identifying regions of interest for both speech types by fMRI

Significant ROIs for both types of speech were identified by T-tests on

fMRI (Fig. 2a) and treated as the spatial constraints for the subsequent analysis of EEG dipole source localization (as shown in Fig. 1b). For OS, we detected a list of typical speech regions (Price, 2010, 2012). For CS, the bilateral SMA was the only significant region. The SMA has been identified in previous CS studies, with some hypothesized functions for the planning and controlling of speech motor (Hertrich et al., 2016), modulation effect towards auditory areas (Linden et al., 2011), and a sign of high neural demand of attention and planning (Shergill et al., 2002). Because BOLD signal from fMRI was an indirect estimation of neural activity with lower temporal resolution, it was not surprising that the BOLD changes were observed only in the SMA but not across a broader functional network during CS. Therefore, we further deepened the exploration by the subsequent multi-modal analysis.

4.2. Spatiotemporal dynamic differences between covert speech and overt speech based on fMRI-informed EEG source localization

The fMRI-informed EEG source localization method integrated the spatiotemporal features of both CS and OS. Three ROIs activated during CS were detected: The left PCG, left SMA, and left putamen (Fig. 3). These ROIs were found to be directly connected in neuroanatomy studies as the major motor-related neural circuit through basal ganglia (Alexander et al., 1986). This circuit was also in line with a motor imagery fMRI study (Makary et al., 2017), which detected higher activations in the circuit during motor imagery than during motor execution.

Furthermore, we explored the spatiotemporal dynamics that could hardly be detected by single modalities with limited resolution like EEG or fMRI. For ECoG, although it offers integrated spatiotemporal resolution, it is invasive and can hardly measure deeper signals from the subcortical regions like the putamen (Parvizi & Kastner, 2018).

In the view of event-locked EEG activation (middle column in Fig. 3), CS had weaker and shorter activations in the left PCG and the left SMA

compared with OS. However, we found that CS was characterized as the earlier event-locked activation in the left putamen (peaking at 262 ms versus 1170 ms), and this difference was still valid after non-parametric re-verifications of bootstrapping and permutation tests (Supplementary Material 5). This activation time of 262 ms in the left putamen was classified into the "lemma retrieval stage" in previous reviews of the temporal dynamics speech production (Indefrey, 2011; Indefrey and Levelt, 2004), regarding as the early brain abstraction stage of word, while in contrast, activation peaks in other regions in this study were all classified into the stage of "phonetic encoding" (onset at ~455 ms), or even the final stage of articulation (onset at ~600 ms). These differences suggest that the early activation of the left putamen might represent the initial temporal stage in the neural dynamics of CS. Our results also revealed a temporal series of neural activity during CS, beginning in the left putamen and subsequently in the left PCG and the left SMA. This sequence of activations partially aligns with the proposed neural model of proactive inhibition (Aron, 2011), which is thought to be associated with CS (Nalborczyk et al., 2022). We also noticed that the left putamen played a key role in the early modulation of motor in previous studies. For example, some epileptic symptoms were interpreted as the dysfunction of putamen during the early regulation of movements (Bartolini et al., 2014; Moeller et al., 2014). Moreover, a neuronal projection analysis in monkeys found that the activation in the putamen was earlier than in the motor cortex and actual movement, meaning an action selection process occurs in the basal ganglia including the putamen (Bronfeld et al., 2011). After combining previous findings with ours, we propose that this dynamic pattern, induced by the early activity in left putamen, might lead to inhibitive modulation during the early stage of speech production, which is a key factor that transforming OS to CS.

For the power spectrum analysis, we consistently detected the larger power in OS than CS in most of the ROIs, which was in line with the previous review (Perrone-Bertolotti et al., 2014), implying apparently weaker neural activity strengths related to motor execution and potential auditory feedback during CS (Oppenheim and Dell, 2010).

The topo-plot in the left putamen revealed stronger differences between CS and OS, compared with the topo-plots of left PCG and SMA. Despite topo-plots providing limited spatial details, this distinction in the left putamen can imply stronger difference in the neural patterns between CS and OS.

4.3. Functional connectivity difference between covert speech and overt speech based on fMRI

Although fMRI-informed EEG source localization has spatiotemporal resolution advantage, it can not detect enough sources data across all ROIs. Therefore, to achieve a comprehensive analysis of functional network, we choice to analyze FC based on the typical fMRI pipeline.

To better interpret the networks, we performed the unsupervised clustering method for identifying the communities (clusters) of regions that had strong intra FC. Notably, CS and OS networks yielded the same cluster structures by this data-driven method, supporting the previous hypothesis that CS and OS partially shared similar functional networks (Alderson-Day and Fernyhough, 2015).

Our FC analysis further emphasized the role of the left putamen as the only hub node in both FC networks during CS and OS (Fig. 4), regarded as a stable connector that consistently facilitating dynamic interactions with multiple language networks (Fedorenko and Thompson-Schill, 2014). As a hub node, the left putamen bridged the left and right clusters (Fig. 4c), aligning with previous findings (Vigneau et al., 2011) about the coactivations in the left putamen towards both hemispheres, associated with both direct and high-order language functions. The lateralization distribution of clusters also justified our inclusion criteria with only male participants for a better data homogeneity, because the sex difference in language lateralization was commonly acknowledged (Sato, 2020; Shaywitz et al., 1995; Zaidi,

2010; Hirnstein et al., 2019).

After quantifying the difference in FC strength, we only detected the lower FC in CS than OS between the left putamen and the left cluster (Fig. 4d). The coactivation between the left putamen and the left cluster has been detected in previous speech studies (Viñas-Guasch and Wu, 2017), which is believed to be directly related to language processing. This FC difference only at the dominant hemisphere is also in line with the previous fMRI-lateralization study between CS and OS (Partovi et al., 2012). The putamen is also observed with achieving motor inhibition by the variance of FC from a Go-NoGo fMRI study (Akkermans et al., 2018) and in a reading-aloud task (Seghier and Price, 2010). As the inhibition of motor function in CS is still debatable (Perrone-Bertolotti et al., 2014), our FC findings provided direct spatiotemporal evidence that the left putamen, as a hub node among the functional network, achieved the inhibition of motor execution by influencing the downstream FC towards the speech-related network, especially at the dominant hemisphere.

4.4. Verifying the effect pathways between multi-modal features using path analysis and proposing the potential mechanisms of CS

We performed PA as a statistical tool to investigate the effect pathways between multi-modal features underlying CS. As a hypothesis-based model, we have explained the data-driven steps of constructing the hypothesized PA model (Fig. 5) in Supplementary Material 8. We constructed PA model by data-driven manner because it is more sensitive than model-driven for detecting relationships between features (Marrelec et al., 2007; Warren et al., 2021), thereby facilitating a comprehensive understanding of the neural mechanisms in CS.

In summary, our model, with a focus on the left putamen, identified both direct and indirect pathways influencing the transformation between CS and OS: 1. Late event-locked EEG activation demonstrated a direct positive association with OS. 2. Conversely, early EEG activations did not show a direct association with the label of CS/OS. Instead, these early activations indirectly contribute to CS through a mediating effect by reduce FC between the left putamen and the left cluster.

Here, by combining the EEG dynamics, FC, and PA results, we propose a potential mechanism interpretation of both pathways in the PA model: the left putamen can be involved in two distinct phases during the neural process of speech. In the early phase (peak at 262 ms), event-locked activation occurs earlier during CS than during OS, involving abstract language processing and inhibition of speech motor functions. This is mediated through the inhibition of FC strength towards the left cluster. In the late phase (peak at 1170 ms), particularly for OS, it might be more involved in the auditory feedback after articulation (Kearney and Guenther, 2019), as the putamen was identified in the related neural circuits (Jürgens, 2009). This was inferred because the late event-locked activation (1170 ms) in the left putamen was only detected during OS but not during CS, which occurred later than the normal reaction time for articulation (~600 ms) (Indefrey, 2011; Indefrey and Levelt, 2004; Jarick and Jones, 2009). It was also later than the activations in most other ROIs in this study (as observed in Fig. 3 and Supplementary Material 4). Additionally, we further double-checked the reliability of the pathways by permutation tests, meaning that the proposed pathways in the model in Fig. 5 fit the real data well, which significantly outperformed the randomly permuted data.

4.5. The left putamen in self-regulation, speech dysfunctions, and BCI

Our CS findings might explain, from a neural mechanistic perspective, why CS plays such a critical role in self-regulation, as neural inhibitory pathways are considered to be the key to self-regulation (Schel et al., 2014). The involvement of the left putamen in the rapid initiation and efficient execution of inhibitory control processes might be essential for successful self-regulation (Jahanshahi et al., 2015).

Based on this hypothesis, our EEG-fMRI findings offer insights into

potential spatial targets and temporal windows for future neuro-linguistic and self-regulation-related behavioral therapies, particularly for deficits in self-regulation and language. For example, these findings suggest future targeted interventions focusing on CS pathways and reinforcing early activation of key regions, which might enhance inhibitory control and facilitate flexible brain FC.

Our findings of CS are complementary with many previous speech-related dysfunctions, implying the shared mechanisms related to the left putamen. Therefore, The CS mechanisms that we proposed in this study might offer the potential treatment for these speech disorders. For example, for auditory verbal hallucination, previous study detected hyper FC between the left putamen and the Wernicke's area (which refers to the left superior temporal gyrus (L-STG) within the left cluster in Fig. 4) (Hoffman and Hampson, 2012). For stuttering speakers, some fMRI researches observed higher activation (Chang et al., 2009), influenced FC (Lu et al., 2010; Yang et al., 2016), and also larger gray matter volume (Lu et al., 2010) in the left putamen. For speech impairment in Parkinson's disease, a previous study detected an influenced strength of FC related to the left putamen and cortical speech regions (Manes et al., 2018) and lesions in the left putamen during aphasia (Barat et al., 1981). For the second language speakers (especially for non-fluent speakers), their left putamen has higher cerebral blood flow by the H_2O^{15} -PET scan (Klein et al., 1994) and larger gray matter volume by MRI (Abutalebi et al., 2013). Higher functional activity by fMRI was observed in the left putamen when they were not highly proficient in a second language (Abutalebi et al., 2013; Liu et al., 2010).

Other than speech disorders, the choice of ROIs and frequencies for decoding CS brain signals into content is important but debatable in the BCI field (Proix et al., 2022), and our study suggests focusing on the brain dynamics between the regions of the left PCG, SMA and putamen, and the functional interactions between the left putamen and speech-related regions, which might be helpful in extracting the CS content from neural measurements like EEG.

4.6. Limitations

1. The inclusion criteria are limited to only male participants to increase sample homogeneity, given plenty of reports had emphasized the gender differences in the brain production of speech (Sato, 2020; Shaywitz et al., 1995; Zaidi, 2010; Hirnstein et al., 2019). Our results also showed strong brain lateralization during speech, which was also acknowledged to be highly different between genders. This limitation can possibly make the findings less generalizable to females.

2. As for the design of the experiment, implementation of speech can have multiple ways: such as word repetition, naming, word generation, reading, and rhyme judgement, which can yield different neural patterns. In our study, we designed a five-word repetition task to focus on the direct effect of CS. The features might be shifted if the experimental manner of speech is changed. We did not try to classify the neural difference between words, as this refers to the speech decoding study in BCI that is beyond our research objective of neural mechanism.

3. The use of fMRI-informed EEG offers good spatiotemporal resolution, but as an optimization analysis, it also presents a potential limitation of higher variability of the results. We should also consider that participants could occasionally become absent-minded, miss some of the speech trials, while during CS trials we could not spot the misses. Based on these two reasons, we could reasonably expect the limitation that not all of the 32 participants contained valid EEG dipole sources within the speech-related ROI, and we calculated the results based on valid EEG dipole sources, inspired by a previous EEG-fMRI study (Timmermann et al., 2023). However, it is noteworthy that our results were still significant based on substantial data amount of thousands of repeated EEG trials and re-verified by bootstrapping and permutation tests. Although EEG-fMRI can substantially improve the spatial precision of neural activity (Michel and Brunet, 2019), decoding EEG-based brain activity from deep subcortical regions like putamen would benefit from

supplementary confirmation through methodologies such as invasive electrophysiological recordings in the future.

4. We designed a 1.5-second trial, which limits the interpretation of our results by omitting information from later stages of cognitive processes for CS and OS. This decision was a necessary compromise to balance the need for a sufficient number of repeated trials and the desire for longer trials within the constraints of the total experiment time. Despite this, the 1.5-second trial was proved adequate for detecting significant neural distinctions between CS and OS, highlighting its effectiveness in capturing early brain dynamics. Finally, we suggest that future research explores extended time windows for a more comprehensive understanding of CS.

CRediT authorship contribution statement

Wei Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Muyun Jiang:** Writing – review & editing, Software, Resources, Methodology, Investigation, Data curation, Conceptualization. **Kok Ann Colin Teo:** Writing – review & editing, Resources, Project administration, Methodology, Funding acquisition, Data curation, Conceptualization. **Raghavan Bhuvanankantham:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Conceptualization. **LaiGuan Fong:** Writing – review & editing, Data curation, Conceptualization. **Wei Khang Jeremy Sim:** Writing – review & editing, Software, Methodology, Conceptualization. **Zhiwei Guo:** Writing – review & editing, Conceptualization. **Chuan Huat Vince Foo:** Writing – review & editing, Supervision, Conceptualization. **Rong Hui Jonathan Chua:** Writing – review & editing, Conceptualization. **Parasuraman Padmanabhan:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Victoria Leong:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization. **Jia Lu:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Balázs Gulyás:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization. **Cuntai Guan:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

No conflict of interest exists in the submission of this manuscript, and the manuscript is approved by all authors for publication.

Data availability

The data that has been used is confidential.

Ethical statement

The study was in accordance with the Declaration of Helsinki and was conducted in Cognitive Neuroimaging Centre (CoNIC) in Nanyang Technological University (NTU) in Singapore. All the participants provided written informed consent. Ethics committee approval was obtained from the Institutional Review Board (IRB) of NTU, with the approval number: IRB-2022-040.

Fundings

W.Z. and V.L. are supported by the RIE2025 Human Potential Programme Prenatal/Early Childhood Grant (H22P0M0002), administered by A*STAR. This work was supported by funding from the Cognitive Neuroimaging Centre, NTU Shared Research Facility (D821/CoNIC),

Ageing Research Institute for Society and Education (ARISE/2017/16) and DSO National Laboratories (DSOCL21193).

Acknowledgements

We thank all the participants for their cooperation during the experiment.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2024.120629.

References

- Abebe, T.H., 2019. The Derivation and choice of appropriate test statistic (z, t, f and chi-square test) in research methodology. *J. Math. Lett* 5 (3), 33–40. https://www.academia.edu/download/65289903/The_Derivation_and_Choice_of_Appropriate_Test_Statistic_Z.pdf.
- Abo, M., Kakuda, W., Watanabe, M., Morooka, A., Kawakami, K., Senoo, A., 2012. Effectiveness of low-frequency rTMS and intensive speech therapy in poststroke patients with aphasia: A pilot study based on evaluation by fMRI in relation to type of aphasia. *Eur. Neurol.* 68 (4), 199–208. <https://karger.com/ene/article-abstract/68/4/199/124986>.
- Abutalebi, J., Della Rosa, P.A., Gonzaga, A.K.C., Keim, R., Costa, A., Perani, D., 2013. The role of the left putamen in multilingual language production. *Brain Lang.* 125 (3), 307–315.
- Akkermans, S.E.A., Luijten, M., van Rooij, D., Franken, I.H.A., Buitelaar, J.K., 2018. Putamen functional connectivity during inhibitory control in smokers and non-smokers. *Addict. Biol.* 23 (1), 359–368. <https://doi.org/10.1111/adb.12482>.
- Albein-Urios, N., Youssef, G., Klas, A., Enticott, P.G., 2021. Inner speech moderates the relationship between autism spectrum traits and emotion regulation. *J. Autism. Dev. Disord.* 51, 3322–3330.
- Alderson-Day, B., Fernyhough, C., 2015. Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychol. Bull.* 141 (5), 931–965. <https://doi.org/10.1037/bul0000021>.
- Alexander, G.E., DeLong, M.R., Strick, P.L., 1986. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annu. Rev. Neurosci.* 9 (1), 357–381.
- Androu, C., Frielinghaus, H., Rauh, J., Mußmann, M., Vauth, S., Braun, P., Leicht, G., Mulert, C., 2017. Theta and high-beta networks for feedback processing: A simultaneous EEG–fMRI study in healthy male subjects. *Transl. Psychiatry* 7 (1), e1016. <https://doi.org/10.1038/tp.2016.287>.
- Aron, A.R., 2011. From reactive to proactive and selective control: Developing a richer model for stopping inappropriate responses. *Biol. Psychiatry* 69 (12), e55–e68. <https://doi.org/10.1016/j.biopsych.2010.07.024>.
- Ashburner, J., Barnes, G., Chen, C.-C., Daunizeau, J., Flandin, G., Friston, K., Kiebel, S., Kilner, J., Litvak, V., Moran, R., 2014. SPM12 manual. Wellcome Trust Centre for Neuroimaging 2464 (4).
- Bak, N., Glenthøj, B.Y., Rostrup, E., Larsson, H.B., Oranje, B., 2011. Source localization of sensory gating: A combined EEG and fMRI study in healthy volunteers. *Neuroimage* 54 (4), 2711–2718. <https://doi.org/10.1016/j.neuroimage.2010.11.039>.
- Barat, M., Mazaux, J.M., Bioulac, B., Giroire, J.M., Vital, C., Arné, L., 1981. Aphasic-type language disorders associated with lesions of the putamen and caudate nucleus: Clinicopathological findings in one case (author's transl). *Rev. Neurol. (Paris)* 137 (5), 343–356. <https://europepmc.org/article/med/6170104>.
- Barber, L., Reniers, R., Uthegrove, R., 2021. A review of functional and structural neuroimaging studies to investigate the inner speech model of auditory verbal hallucinations in schizophrenia. *Transl. Psychiatry* 11 (1), 582. <https://doi.org/10.1038/s41398-021-01670-7>.
- Barkley, R.A., 1997. Attention-deficit/hyperactivity disorder, self-regulation, and time: Toward a more comprehensive theory. *J. Dev. Behav. Pediatr.* 18 (4), 271–279. https://journals.lww.com/jrnl/abstract/1997/08000/Attention_Deficit_Hyperactivity_Disorder.9.aspx.
- Bartolini, E., Pesaresi, I., Fabbri, S., Cecchi, P., Giorgi, F.S., Sartucci, F., Bonucelli, U., Cosottini, M., 2014. Abnormal response to photic stimulation in Juvenile Myoclonic Epilepsy: An EEG-fMRI study. *Epilepsia* 55 (7), 1038–1047.
- Basho, S., Palmer, E.D., Rubio, M.A., Wulfeck, B., Müller, R.-A., 2007. Effects of generation mode in fMRI adaptations of semantic fluency: Paced production and overt speech. *Neuropsychologia* 45 (8), 1697–1706.
- Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. *J. Stat. Mech.: Theory Exp.* 2008 (10), P10008.
- BrainVision Analyzer (Version 2.2.2). (2021). [Computer software]. Brain Products GmbH.
- Brocklehurst, P.H., Corley, M., 2011. Investigating the inner speech of people who stutter: Evidence for (and against) the covert repair hypothesis. *J. Commun. Disord.* 44 (2), 246–260.
- Bronfeld, M., Belevsky, K., Bar-Gad, I., 2011. Spatial and temporal properties of tic-related neuronal activity in the cortico-basal ganglia loop. *J. Neurosci.: Off. J. Soc. Neurosci.* 31 (24), 8713–8721. <https://doi.org/10.1523/JNEUROSCI.0195-11.2011>.
- Burle, B., Spieser, L., Roger, C., Casini, L., Hasbroucq, T., Vidal, F., 2015. Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *Int. J. Psychophysiol.* 97 (3), 210–220. <https://www.sciencedirect.com/science/article/pii/S0167876015001865>.
- Chang, S.-E., Kenney, M.K., Loucks, T.M., Ludlow, C.L., 2009. Brain activation abnormalities during speech and non-speech in stuttering speakers. *Neuroimage* 46 (1), 201–212.
- Cohen, J., 2013. *Statistical Power Analysis For the Behavioral Sciences*. Academic press. Community Detection Toolbox—File Exchange—MATLAB Central. (n.d.). Retrieved April 17, 2024, from <https://www.mathworks.com/matlabcentral/fileexchange/45867-community-detection-toolbox>.
- Cooney, C., Folli, R., Coyle, D., 2018. Neurolinguistics research advancing development of a direct-speech brain-computer interface. *iScience* 8, 103–125. [https://www.cell.com/iscience/pdf/S2589-0042\(18\)30151-2.pdf](https://www.cell.com/iscience/pdf/S2589-0042(18)30151-2.pdf).
- Day, K.L., Smith, C.L., 2013. Understanding the role of private speech in children's emotion regulation. *Early. Child Res. Q.* 28 (2), 405–414. <https://www.sciencedirect.com/science/article/pii/S0885200612001019>.
- Dehghani, A., Soltanian-Zadeh, H., Hossein-Zadeh, G.-A., 2023. Probing fMRI brain connectivity and activity changes during emotion regulation by EEG neurofeedback. *Front. Hum. Neurosci.* 16, 988890. <https://www.frontiersin.org/articles/10.3389/fnhum.2022.988890/full>.
- 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 (1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- Duffy, J.R., Watt, J., Duffy, R.J., 1981. Path Analysis. *J. Speech, Lang., Hear. Res.* 24 (4), 474–490. <https://doi.org/10.1044/jshr.2404.474>.
- Fedorenko, E., Thompson-Schill, S.L., 2014. Reworking the language network. *Trends. Cogn. Sci.* 18 (3), 120–126. [https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613\(13\)00288-X](https://www.cell.com/trends/cognitive-sciences/fulltext/S1364-6613(13)00288-X).
- Fernyhough, C., Watson, A., Bernini, M., Moseley, P., Alderson-Day, B., 2019. Imaginary companions, inner speech, and auditory verbal hallucinations: What are the relations? *Front. Psychol.* 10, 1665.
- Geva, S., Jones, P.S., Crinion, J.T., Price, C.J., Baron, J.-C., Warburton, E.A., 2011. The neural correlates of inner speech defined by voxel-based lesion–symptom mapping. *Brain* 134 (10), 3071–3082. <https://academic.oup.com/brain/article-abstract/134/10/3071/323105>.
- Golkowski, D., Merz, K., Mlynarcik, C., Kiel, T., Schorr, B., Lopez-Rolon, A., Lukas, M., Jordan, D., Bender, A., Ilg, R., 2017. Simultaneous EEG–PET–fMRI measurements in disorders of consciousness: An exploratory study on diagnosis and prognosis. *J. Neurol.* 264 (9), 1986–1995. <https://doi.org/10.1007/s00415-017-8591-z>.
- Gonzalez-Lopez, J.A., Gomez-Alanis, A., Doñas, J.M.M., Pérez-Córdoba, J.L., Gomez, A.M., 2020. Silent speech interfaces for speech restoration: A Review. *IEE Access* 8, 177995–178021.
- Goto, T., Hirata, M., Umekawa, Y., Yanagisawa, T., Shayne, M., Saitoh, Y., Kishima, H., Yorifuji, S., Yoshimine, T., 2011. Frequency-dependent spatiotemporal distribution of cerebral oscillatory changes during silent reading: A magnetoencephalographic group analysis. *Neuroimage* 54 (1), 560–567. <https://doi.org/10.1016/j.neuroimage.2010.08.023>.
- Hertrich, I., Dietrich, S., Ackermann, H., 2016. The role of the supplementary motor area for speech and language processing. *Neurosci. Biobehav. Rev.* 68, 602–610. <https://doi.org/10.1016/j.neubiorev.2016.06.030>.
- Hirnstein, M., Hugdahl, K., Hausmann, M., 2019. Cognitive sex differences and hemispheric asymmetry: A critical review of 40 years of research. *Laterality: Asymmet. Body, Brain Cognit.* 24 (2), 204–252.
- Hoffman, R.E., Hampson, M., 2012. Functional connectivity studies of patients with auditory verbal hallucinations. *Front. Hum. Neurosci.* 6, 6.
- Huster, R.J., Debener, S., Eichele, T., Herrmann, C.S., 2012. Methods for simultaneous EEG–fMRI: An introductory review. *J. Neurosci.* 32 (18), 6053–6060. <https://doi.org/10.1523/JNEUROSCI.0447-12.2012>.
- Indefrey, P., 2011. The Spatial and Temporal Signatures of Word Production Components: A Critical Update. *Front. Psychol.* 2. <https://www.frontiersin.org/articles/10.3389/fpsyg.2011.00255>.
- Indefrey, P., Levelt, W.J., 2004. The spatial and temporal signatures of word production components. *Cognition* 92 (1–2), 101–144. <https://www.sciencedirect.com/science/article/pii/S0010027703002294>.
- Jahanshahi, M., Obeso, I., Rothwell, J.C., Obeso, J.A., 2015. A fronto–striato–subthalamic–pallidal network for goal-directed and habitual inhibition. *Nat. Rev. Neurosci.* 16 (12), 719–732. <https://www.nature.com/articles/nrn4038>.
- Jarick, M., Jones, J.A., 2009. Effects of seeing and hearing speech on speech production: A response time study. *Exp. Brain Res.* 195 (2), 175–182. <https://doi.org/10.1007/s00221-009-1765-x>.
- Jürgens, U., 2009. The neural control of vocalization in mammals: A review. *J. Voice* 23 (1), 1–10.
- Kearney, E., Guenther, F.H., 2019. Articulating: The neural mechanisms of speech production. *Lang. Cogn. Neurosci.* 34 (9), 1214–1229. <https://doi.org/10.1080/23273798.2019.1589541>.
- Kim, S., Richter, W., Uğurbil, K., 1997. Limitations of temporal resolution in functional MRI. *Magn. Reson. Med.* 37 (4), 631–636. <https://doi.org/10.1002/mrm.1910370427>.
- Klein, D., Zatorre, R.J., Milner, B., Meyer, E., Evans, A.C., 1994. Left putamen activation when speaking a second language: Evidence from PET. *Neuroreport* 5 (17), 2295–2297.
- Langland-Hassan, P., Faries, F.R., Richardson, M.J., Dietz, A., 2015. Inner speech deficits in people with aphasia. *Front. Psychol.* 6. <https://www.frontiersin.org/articles/10.3389/fpsyg.2015.00528>.

- Lidstone, J.S., Meins, E., Fernyhough, C., 2010. The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual task paradigm. *J. Exp. Child Psychol.* 107 (4), 438–451.
- Linden, D.E.J., Thornton, K., Kuswanto, C.N., Johnston, S.J., Van De Ven, V., Jackson, M. C., 2011. The Brain's Voices: Comparing Nonclinical Auditory Hallucinations and Imagery. *Cerebral Cortex* 21 (2), 330–337. <https://doi.org/10.1093/cercor/bhq097>.
- Liu, H., Hu, Z., Guo, T., Peng, D., 2010. Speaking words in two languages with one brain: Neural overlap and dissociation. *Brain Res.* 1316, 75–82.
- Lopes da Silva, F., 2013. EEG and MEG: relevance to neuroscience. *Neuron* 80 (5), 1112–1128. <https://doi.org/10.1016/j.neuron.2013.10.017>.
- Lu, C., Peng, D., Chen, C., Ning, N., Ding, G., Li, K., Yang, Y., Lin, C., 2010. Altered effective connectivity and anomalous anatomy in the basal ganglia-thalamocortical circuit of stuttering speakers. *Cortex* 46 (1), 49–67. <https://doi.org/10.1016/j.cortex.2009.02.017>.
- Makary, M.M., Eun, S., Park, K., 2017. Greater corticostriatal activation associated with facial motor imagery compared with motor execution: A functional MRI study. *Neuroreport* 28 (10), 610–617. <https://doi.org/10.1097/WNR.0000000000000809>.
- Manes, J.L., Tjaden, K., Parrish, T., Simuni, T., Roberts, A., Greenlee, J.D., Corcos, D.M., Kurani, A.S., 2018. Altered resting-state functional connectivity of the putamen and internal globus pallidus is related to speech impairment in Parkinson's disease. *Brain Behav.* 8 (9), e01073. <https://doi.org/10.1002/brb3.1073>.
- Marrelec, G., Horwitz, B., Kim, J., Pélégrini-Issac, M., Benali, H., Doyon, J., 2007. Using partial correlation to enhance structural equation modeling of functional MRI data. *Magn. Reson. Imaging* 25 (8), 1181–1189. <https://doi.org/10.1016/j.mri.2007.02.012>.
- Martin, S., Brunner, P., Holdgraf, C., Heinze, H.-J., Crone, N.E., Rieger, J., Schalk, G., Knight, R.T., Pasley, B.N., 2014. Decoding spectrotemporal features of overt and covert speech from the human cortex. *Front. Neuroeng.* 7. <https://www.frontiersin.org/articles/10.3389/fneng.2014.00014>.
- Martínez-Manrique, F., Vicente, A., 2015. The activity view of inner speech. *Front. Psychol.* 6, 232.
- Marvel, C.L., Desmond, J.E., 2012. From storage to manipulation: How the neural correlates of verbal working memory reflect varying demands on inner speech. *Brain Lang.* 120 (1), 42–51.
- Meyer, L., Obleser, J., Kiebel, S.J., Friederici, A.D., 2012. Spatiotemporal dynamics of argument retrieval and reordering: An fMRI and EEG study on sentence processing. *Front. Psychol.* 3, 523.
- Michel, C.M., Brunet, D., 2019. EEG source imaging: A practical review of the analysis steps. *Front. Neurosci.* 10, 325. <https://www.frontiersin.org/articles/10.3389/fnec.2019.00325/full>.
- Moeller, F., Groening, K., Moehring, J., Muhle, H., Wolff, S., Jansen, O., Stephani, U., Siniatchkin, M., 2014. EEG-fMRI in myoclonic astatic epilepsy (Doose syndrome). *Neurology.* 82 (17), 1508–1513.
- Musso, M., Hübner, D., Schwarzkopf, S., Bernodsson, M., LeVan, P., Weiller, C., Tangermann, M., 2022. Aphasia recovery by language training using a brain-computer interface: A proof-of-concept study. *Brain Commun.* 4 (1), fcac008. <https://academic.oup.com/braincomms/article-abstract/4/1/fcac008/6524563>.
- Muthukumaraswamy, S., 2013. High-frequency brain activity and muscle artifacts in MEG/EEG: A review and recommendations. *Front. Hum. Neurosci.* 7. <https://www.frontiersin.org/articles/10.3389/fnhum.2013.00138>.
- Nalborczyk, L., Debarnot, U., Longcamp, M., Guillot, A., Alario, F.-X., 2022. The role of motor inhibition during overt speech production. *Front. Hum. Neurosci.* 16, 804832. <https://doi.org/10.3389/fnhum.2022.804832>.
- Oostendorp, T.F., Van Oosterom, A., 1989. Source parameter estimation in inhomogeneous volume conductors of arbitrary shape. *IEEE Trans. Biomed. Eng.* 36 (3), 382–391.
- Oostenfeld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, 1–9.
- Oostenfeld, R., Praamstra, P., Stegeman, D.F., Van Oosterom, A., 2001. Overlap of attention and movement-related activity in lateralized event-related brain potentials. *Clin. Neurophysiol.* 112 (3), 477–484.
- Oppenheim, G.M., Dell, G.S., 2010. Motor movement matters: The flexible abstractness of inner speech. *Mem. Cognit.* 38 (8), 1147–1160.
- Partovi, S., Konrad, F., Karimi, S., Rengier, F., Lyo, J.K., Zipp, L., Nennig, E., Stippich, C., 2012. Effects of covert and overt paradigms in clinical language fMRI. *Acad. Radiol.* 19 (5), 518–525. <https://doi.org/10.1016/j.acra.2011.12.017>.
- Parvizi, J., Kastner, S., 2018. Human intracranial EEG: Promises and limitations. *Nat. Neurosci.* 21 (4), 474. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6476542/>.
- Pei, X., Leuthardt, E.C., Gaona, C.M., Brunner, P., Wolpaw, J.R., Schalk, G., 2011. Spatiotemporal dynamics of electrocorticographic high gamma activity during overt and covert word repetition. *Neuroimage* 54 (4), 2960–2972. <https://doi.org/10.1016/j.neuroimage.2010.10.029>.
- Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J.K., 2019. PsychoPy2: Experiments in behavior made easy. *Behav. Res. Methods* 51 (1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>.
- Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciú, M., Lœvenbruck, H., 2014. What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behav. Brain Res.* 261, 220–239. <https://doi.org/10.1016/j.bbr.2013.12.034>.
- Pion-Tonachini, L., Kreutz-Delgado, K., Makeig, S., 2019. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage* 198, 181–197.
- Power, J.D., Schlaggar, B.L., Lessov-Schlaggar, C.N., Petersen, S.E., 2013. Evidence for hubs in human functional brain networks. *Neuron* 79 (4), 798–813.
- Price, C.J., 2010. The anatomy of language: A review of 100 fMRI studies published in 2009. *Ann. N. Y. Acad. Sci.* 1191 (1), 62–88.
- Price, C.J., 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* 62 (2), 816–847.
- Proix, T., Delgado Saa, J., Christen, A., Martin, S., Pasley, B.N., Knight, R.T., Tian, X., Poeppel, D., Doyle, W.K., Devinsky, O., Arnal, L.H., Mégevand, P., Giraud, A.-L., 2022. Imagined speech can be decoded from low- and cross-frequency intracranial EEG features. *Nat. Commun.* 13 (1), 48. <https://doi.org/10.1038/s41467-021-27725-3>.
- Richter, W., Richter, M., 2003. The shape of the fMRI BOLD response in children and adults changes systematically with age. *Neuroimage* 20 (2), 1122–1131. <https://www.sciencedirect.com/science/article/pii/S1053811903003471>.
- Rosseel, Y., 2012. lavaan: An R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36. <https://www.jstatsoft.org/article/view/v48i02>.
- Sato, M.H., 2020. The neurobiology of sex differences during language processing in healthy adults: A systematic review and a meta-analysis. *Neuropsychologia* 140, 107404.
- Schel, M.A., Scheres, A., Crone, E.A., 2014. New perspectives on self-control development: Highlighting the role of intentional inhibition. *Neuropsychologia* 65, 236–246. <https://www.sciencedirect.com/science/article/pii/S0028393214002851>.
- Schermelleh-Engel, K., Moosbrugger, H., Müller, H., 2003. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods Psychol. Res. Online* 8 (2), 23–74.
- Schrooten, M., Vandenberghe, R., Peeters, R., Dupont, P., 2018. Quantitative Analyses Help in Choosing Between Simultaneous vs. Separate EEG and fMRI. *Front. Neurosci.* 12, 1009. <https://doi.org/10.3389/fnins.2018.01009>.
- Seghier, M.L., Price, C.J., 2010. Reading aloud boosts connectivity through the putamen. *Cerebral Cortex* 20 (3), 570–582.
- Shaywitz, B.A., Shaywitz, S.E., Pugh, K.R., Constable, R.T., Skudlarski, P., Fulbright, R. K., Bronen, R.A., Fletcher, J.M., Shankweiler, D.P., Katz, L., 1995. Sex differences in the functional organization of the brain for language. *Nature* 373 (6515), 607–609. <https://www.nature.com/articles/373607a0>.
- Shergill, S.S., Brammer, M.J., Fukuda, R., Bullmore, E., Amaro Jr, E., Murray, R.M., McGuire, P.K., 2002. Modulation of activity in temporal cortex during generation of inner speech. *Hum. Brain Mapp.* 16 (4), 219–227.
- Simistira Liwicki, F., Gupta, V., Saini, R., De, K., Abid, N., Rakesh, S., Wellington, S., Wilson, H., Liwicki, M., Eriksson, J., 2023. Bimodal electroencephalography-functional magnetic resonance imaging dataset for inner-speech recognition. *Sci. Data* 10 (1), 378.
- Stark, B.C., Geva, S., Warburton, E.A., 2017. Inner speech's relationship with overt speech in poststroke aphasia. *J. Speech, Lang., Hear. Res.* 60 (9), 2406–2415. <https://doi.org/10.1044/2017.JSLHR-16-0270>.
- Stephan, F., Saalbach, H., Rossi, S., 2020. The brain differentially prepares inner and overt speech production: Electrophysiological and vascular evidence. *Brain Sci.* 10 (3), 148.
- Stoeckel, L.E., Garrison, K.A., Ghosh, S.S., Wightton, P., Hanlon, C.A., Gilman, J.M., Greer, S., Turk-Browne, N.B., deBettencourt, M.T., Scheinost, D., 2014. Optimizing real time fMRI neurofeedback for therapeutic discovery and development. *NeuroImage: Clin.* 5, 245–255. <https://www.sciencedirect.com/science/article/pii/S2213158214000928>.
- Timmermann, C., Roseman, L., Haridas, S., Rosas, F.E., Luan, L., Kettner, H., Martell, J., Erritzoe, D., Tagliazucchi, E., Pallavicini, C., Girn, M., Alamia, A., Leech, R., Nutt, D. J., Carhart-Harris, R.L., 2023. Human brain effects of DMT assessed via EEG-fMRI. *Proc. Natl. Acad. Sci.* 120 (13), e2218949120. <https://doi.org/10.1073/pnas.2218949120>.
- Tullett, A.M., Inzlicht, M., 2010. The voice of self-control: Blocking the inner voice increases impulsive responding. *Acta Psychol. (Amst)* 135 (2), 252–256. <https://www.sciencedirect.com/science/article/pii/S0001691810001368>.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15 (1), 273–289. <https://doi.org/10.1006/nimg.2001.0978>.
- Ullsperger, M., Von Cramon, D.Y., 2001. Subprocesses of performance monitoring: A dissociation of error processing and response competition revealed by event-related fMRI and ERPs. *Neuroimage* 14 (6), 1387–1401.
- Vigneau, M., Beaucousin, V., Hervé, P.-Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B., Tzourio-Mazoyer, N., 2011. What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing?: Insights from a meta-analysis. *Neuroimage* 54 (1), 577–593. <https://www.sciencedirect.com/science/article/pii/S1053811910010098>.
- Viñas-Guasch, N., Wu, Y.J., 2017. The role of the putamen in language: A meta-analytic connectivity modeling study. *Brain Struct. Funct.* 222, 3991–4004.
- Voigtlaender, V.A., Sandhaeger, F., Hawellek, D.J., Hage, S.R., Siegel, M., 2023. Neural representations of the content and production of human vocalization. *Proc. Natl. Acad. Sci.* 120 (23), e2219310120. <https://doi.org/10.1073/pnas.2219310120>.
- Warren, H.J., Ioachim, G., Powers, J.M., Stroman, P.W., 2021. How fMRI analysis using structural equation modeling techniques can improve our understanding of pain processing in fibromyalgia. *J. Pain Res.* 14, 381–398. <https://doi.org/10.2147/JPR.S290795>.
- Wilson, H., Wellington, S., Liwicki, F.S., Gupta, V., Saini, R., De, K., Abid, N., Rakesh, S., Eriksson, J., Watts, O., 2023. Performance of data-driven inner speech decoding with same-task EEG-fMRI data fusion and bimodal models. *arXiv Preprint arXiv: 2306.10854*.

Yang, Y., Jia, F., Siok, W.T., Tan, L.H., 2016. Altered functional connectivity in persistent developmental stuttering. *Sci. Rep.* 6 (1), 19128. <https://doi.org/10.1038/srep19128>.

Yoshimura, N., Nishimoto, A., Belkacem, A.N., Shin, D., Kambara, H., Hanakawa, T., Koike, Y., 2016. Decoding of covert vowel articulation using electroencephalography

cortical currents. *Front. Neurosci.* 10. <https://www.frontiersin.org/articles/10.3389/fnins.2016.00175>.

Zaidi, Z.F., 2010. Gender differences in human brain: A review. *Open Anatomy J.* 2 (1).