

Functional connectivity between parietal and temporal lobes mediates internal forward models during speech production

Wenjia Zhang^{a,b,c,*}, Fuyin Yang^{b,d}, Xing Tian^{b,c,*}

^a Key Laboratory for Artificial Intelligence and Cognitive Neuroscience of Language, Xi'an International Studies University, Xi'an, China

^b NYU-ECNU Institute of Brain and Cognitive Science at NYU Shanghai, Shanghai, China

^c Division of Arts and Sciences, New York University Shanghai, Shanghai, China

^d Shanghai Key Laboratory of Brain Functional Genomics (Ministry of Education), School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

ARTICLE INFO

Keywords:

Internal forward model
Speech production
Motor simulation
Sensory estimation
Somatosensory
Auditory
Functional connectivity
Inferior parietal lobe (IPL)
Superior temporal gyrus (STG)
Inferior frontal gyrus (IFG)

ABSTRACT

Internal forward models hypothesize functional links between motor and sensory systems for predicting the consequences of actions. Recently, the cascaded theory proposes that somatosensory estimation in the inferior parietal lobe (IPL) can be a relay computational structure, converting motor signals into predictions of auditory consequences in a serial processing manner during speech production. The study used fMRI with functional connectivity (FC) analyses to investigate the proposed cascaded processes using three speech tasks: overt articulation (OA), silent articulation (SA) and imagined articulation (IA). The FC results showed that connectivity between aIPL and STG was increased in OA compared with SA, suggesting that the relationship between somatosensory and auditory estimations can be modulated by speech tasks. Moreover, stronger connectivity between IFGoper and pIPL, and between pIPL and STG were observed in SA and IA compared with OA. These results are consistent with a cascaded process in the internal forward models.

1. Introduction

Internal forward models have been theorized and implicated as an important component in motor control (Wolpert & Ghahramani, 2000). Speech production, one of the most complicated situations of motor control, has been hypothesized to utilize the internal forward model for efficient online adjustment. Specifically, the internal forward model facilitates speech monitoring, error detection, and correction in speech motor control by providing and contrasting the estimates of speech consequences with feedback (Guenther, Ghosh, & Tourville, 2006; Hickok, 2012; Tian & Poeppel, 2012). The computational neuroanatomy and framework of the internal forward models have been demonstrated in major speech production models, including DIVA (Guenther, 1995; Guenther et al., 2006), state feedback control model (Houde & Nagarajan, 2011), hierarchical state feedback control (HSFC) (Hickok, 2012; Hickok, Houde, & Rong, 2011) and dual-stream prediction model (DSPM) (Tian & Poeppel, 2010, 2012; Tian, Zarate, & Poeppel, 2016). Despite the discrepancy in details across models, it is universally accepted that internal forward models in speech production include three systems – motor, somatosensory, and auditory. The motor codes of

speech production are first programmed in the frontal cortices (Guenther et al., 2006; Hickok, 2012; Houde & Chang, 2015; Tian & Poeppel, 2010). Subsequently, a copy of the motor commands (termed efference copy) is used to estimate the somatosensory and auditory consequences of articulation, respectively, in the sensory regions. Our goal is to examine the dynamic relationship between somatosensory and auditory estimation during speech production.

A seminal behavioral study provided convincing evidence suggesting the co-existence of somatosensory and auditory estimations and the interaction between these two estimations during speech production (Lametti, Nasir, & Ostry, 2012). In this study, perturbations in the somatosensory and auditory domains were introduced simultaneously when participants were producing sounds. They found a proportion of participants only compensated for auditory or somatosensory perturbation. However, about a third of participants compensated in both domains, and more importantly, the compensation in auditory and somatosensory domains was negatively correlated. That is, participants who compensated more for the auditory perturbation compensated less for the somatosensory perturbation and vice versa. Moreover, a recent study found compensation was larger in laryngeal perturbation without

* Corresponding authors.

E-mail addresses: zhangwenjia2732@126.com (W. Zhang), xing.tian@nyu.edu (X. Tian).

<https://doi.org/10.1016/j.bandl.2023.105266>

Received 11 September 2022; Received in revised form 4 April 2023; Accepted 15 April 2023

Available online 25 April 2023

0093-934X/© 2023 Elsevier Inc. All rights reserved.

auditory masking than that with auditory masking (Smith, Stepp, Guenther, & Kearney, 2020). These results suggest that both somatosensory and auditory estimations are important components in speech production, as well as possible interaction between the two estimation processes.

Neuroimaging studies offer direct evidence further suggesting the existence of both auditory and somatosensory estimations during speech production, as well as the neural anatomical bases for these two estimations. Speaking suppresses the early auditory responses to auditory feedback (Houde, Nagarajan, Sekihara, & Merzenich, 2002). A series of studies with adaption paradigms found that auditory estimation possesses similar neural representation as auditory perception in primary and associative auditory cortices at the level of phonology, phonetics, and even basic attributes such as intensity (Tian, Ding, Teng, Bai, & Poeppel, 2018; Tian & Poeppel, 2013, 2015). When auditory feedback is online altered, early auditory responses increase as a function of auditory perturbation (Behroozmand & Larson, 2011; Behroozmand, Liu, & Larson, 2011). This evidence supports that the temporal auditory cortices mediate the auditory estimation in the internal forward model. Whereas in the sensorimotor control studies, self-induced touching action suppresses the activity in the parietal somatosensory and supramarginal cortices (Blakemore, Wolpert, & Frith, 1998; Kilteni & Ehrsson, 2020), suggesting these areas mediate somatosensory estimation. In the speech domain, similar areas in the inferior parietal lobe (IPL), including supramarginal gyrus (SMG), parietal operculum (PO), and angular gyrus (AG), have been observed during overt and covert speech tasks (Tian et al., 2016; Zhang, Liu, Wang, & Tian, 2020). These results suggest that the IPL mediates the somatosensory estimation and the temporal auditory cortices mediate the auditory estimation in the internal forward model.

The internal forward models necessitate a distributed network in which motor and sensory regions communicate with each other to transmit information and achieve the computational functions of estimation and prediction (Duffau, Gatignol, Denvil, Lopes, & Capelle, 2003; Tourville, Reilly, & Guenther, 2008). For example, the motor and somatosensory regions need to be “functionally connected” to achieve the somatosensory estimation in the internal forward models. The enhanced estimation in the somatosensory areas would be caused by the strengthened functional connectivity (FC) and correlated activation between the motor and somatosensory regions. That is, the functions of internal forward models during speech production can be better examined by investigating the properties of functional connectivity in the proposed neural network. However, most previous studies examined these two estimation stages from an independent perspective using univariate analyses on activation changes. It is unclear about the relation among the frontal motor, parietal somatosensory, and temporal auditory computational centers of the entire internal forward model in a dynamic and holistic perspective.

The somatosensory estimation in the parietal lobe is a key component to test different hypotheses about internal forward models during speech production and control. The parallel processing account proposes two parallel estimation pathways from the motor to both the somatosensory and auditory regions (Guenther et al., 2006; Houde & Nagarajan, 2011). Whereas, based on previous studies (Chu, Ma, Hang, & Tian, 2022; Hickok, 2012; Poeppel, 2014; Tian et al., 2016), the cascaded processing account implies that the somatosensory may serve as a relay hub in the estimation hierarchy that links between motor and auditory regions. The above-reviewed behavioral study (Lametti et al., 2012) has offered preliminary evidence suggesting the co-existence of parallel and cascaded internal forward models during speech production. That is, when simultaneously perturbed the feedback in the auditory and somatosensory domains, distinct groups of participants adapted in auditory or somatosensory domains, or both domains in a negatively correlated manner. Moreover, a recent neuroimaging study using imagery speech showed the functional connectivity in both direct motor-auditory and cascaded motor-somatosensory-auditory pathways (Chu

et al., 2022). However, whether the cascaded internal forward model exists during overt speech production is unclear.

Moreover, the stages of somatosensory and auditory estimations in the internal forward models are flexible during speech production. For example, the operation of these two stages may be modulated by speech production tasks (Okada, Matchin, & Hickok, 2017; Zhang et al., 2020). Three speech tasks were commonly used to examine the internal forward models in speech production: overt articulation (OA), silent articulation (SA), and imagined articulation (IA). In the OA task, both auditory feedback and somatosensory feedback are available. Therefore, somatosensory and auditory domains may work together to determine speech errors, which would induce increased FC between these two domains. Whereas, in the SA and IA tasks, auditory feedback is absent, which would strengthen the contribution from the somatosensory domain during the speech production task. That speech task demand modulates connectivity between the somatosensory and auditory estimations can provide further evidence for differentiating the hypotheses about parallel and cascaded processes of internal forward models during speech production.

The present study used three speech tasks (OA, SA, and IA) and functional magnetic resonance imaging (fMRI) to systematically investigate the internal forward models during speech production. We first used univariate analyses to reveal the anatomical regions that mediate major computations in the internal forward models – conjunction among all three tasks for frontal motor simulation and parietal somatosensory estimation regions, and the contrast between OA and SA for temporal auditory regions. Then, we conducted FC analyses among these regions and examined how the connectivity was modulated by speech tasks.

We predicted that the connectivity between frontal motor and temporal auditory regions would be stronger in OA than those in SA and IA tasks because the available auditory feedback would weigh the contribution in the auditory more than the somatosensory domain. More importantly, we examined how the connectivity involved in the parietal somatosensory regions would be modulated by tasks. Specifically, both auditory and somatosensory feedbacks are available in the OA task. Therefore, more cognitive demands may be required to connect somatosensory and auditory domains for collaboratively determining the source of errors and controlling speech production. Whereas in the SA and IA tasks, the auditory feedback is absent which would “emphasize” the contribution in the somatosensory domain rather than the interaction between the somatosensory and auditory domains. Therefore, we predicted that if FC between somatosensory and auditory regions was available in all tasks, the connectivity strength between these two regions would be stronger in OA than those in SA and IA tasks. These results would be consistent with a cascaded process in the internal forward models during speech production.

2. Methods

2.1. Participants

Twenty-two volunteers (10 males, mean age: 22.87, SD: 2.69) participated in the fMRI experiment. All participants were right-handed Mandarin Chinese speakers with normal or corrected to normal vision. They reported no history of speech or language disorders. All materials and protocols were approved by the New York University Shanghai Institutional Review Board (IRB). Written informed consent was obtained before the experiment. All participants received monetary incentives for their participation.

2.2. Stimuli and procedure

The stimuli and procedure are similar to that in Zhang et al. (2020).¹ We combined each of the eight consonants (/b/, /p/, /d/, /t/, /w/, /f/, /z/, and /s/) with each of the two vowels (/a/ and /u/) to construct 16 Chinese consonant–vowel syllables (/ba/, /pa/, /da/, /ta/, /wa/, /fa/, /za/, /sa/, /bu/, /pu/, /du/, /tu/, /wu/, /fu/, /zu/ and /su/). All syllables are common in the conversational speech of Mandarin Chinese. Three tasks were used in this study. In the *overt articulation* task (OA), participants were asked to overtly articulate each syllable. In the *silent articulation* task (SA), they were asked to articulate each syllable without phonation. In the *imagined articulation* task (IA), they were asked to imagine speaking each syllable without overtly articulating. Note that participants were required to generate the articulatory movements in the SA task. However, such movements were strongly discouraged in the IA task. Participants were asked to articulate all syllables with the first tone (high) in Chinese for all tasks.

Each trial began with a visual cue that indicated the task and the syllable for the task. A 1200 ms blank, 600 ms green circle, and 600 ms blank were then presented sequentially. Participants were asked to begin responding to a task time-locked to the onset of the green circle. They were asked to articulate the syllable only once at their comfortable speed and duration to ensure the naturalness of articulation. The inter-trial interval (ITI) was randomly chosen from 4440 to 6660 ms (2 to 3 TRs), temporally jittered by 148 ms increments. Each syllable was presented three times in each run. Each run also included six resting trials that were visually cued with the word ‘rest’. The order of trials in each run was randomized. Each participant completed 6 runs, with each task containing two runs². The order of the tasks was presented with the Latin Square Design across participants to control the task order effect. Each participant was asked to conduct a practice session to familiarize themselves with the task before the fMRI experiment.

2.3. MRI scanning

MRI scanning was performed with a Siemens MAGNETOM Prismafit system at East China Normal University. Functional data were acquired using a gradient-echo, echo-planar pulse (EPI) sequence (TR = 2220 ms; TE = 30 ms; 38 slices; field of view = 192 × 192 mm²; 3 × 3 × 3 mm³ voxel size with 0.6 mm interslice gap). We rotated the scanning orientation counter-clockwise about 30° from the AC-PC line to maximize the coverage. High-resolution T1-weighted anatomical images were collected before the functional scan from each participant. Specifically, these images were acquired with a magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence and sagittal slice orientation (176 slices, TR = 1900 ms, TE = 2.53 ms, FOV = 256 mm × 256 mm, flip angle = 9°, voxel size = 1 × 1 × 1 mm³, duration = 4 min 26 s).

2.4. Preprocessing

MRI data were analyzed using SPM12 (<https://www.fil.ion.ucl.ac.uk/spm/>) and customized functions in Matlab. All functional images from each participant were temporally interpolated to the first slice of each volume and spatially realigned to the mean image. The mean functional images were coregistered with the structural image and then segmented. Functional images were then spatially normalized onto the

¹ The research topic of the current study is to investigate the functional connectivity in the proposed motor-to-sensory transformation neural network, which contrasts with the previously published study in which we used the RSA method to examine the neural representation in each of the ROIs in the network.

² Note each task contained two runs in the present study, whereas, each task contained only one run in Zhang et al., 2020.

Montreal Neurological Institute (MNI) space and smoothed with a Gaussian kernel of 6-mm full width at half maximum (FWHM).

2.5. Univariate analyses

Events (including visual cues and articulation tasks) were modeled as sustained boxcar epochs spanning their corresponding duration. Events were convolved with a canonical hemodynamic response function (HRF) implemented in SPM12 and entered as regressors into a general linear model (GLM) for each individual. Each GLM also included head motion regressors and session-wise baseline regressors. The GLM was then estimated using functional data with high pass filtered at 1/128 Hz. Statistical maps of individual-level contrasts were then constructed using the beta estimates of regressors of interest.

For the group-level analyses, we performed voxel-wise t-tests for each pair of task contrasts to systematically examine the task effects and to define ROIs for subsequent FC analyses. To investigate the overlapped activations across tasks, we also conducted conjunction analyses of all three tasks. The correction for multiple comparisons was 0.001 at the voxel level and 0.05 at the cluster level with a family-wise error (FWE) correction. We reported the resultant significance as t-value maps in MNI space.

2.6. FC analyses

For the FC analyses, eight ROIs were created based on the results of univariate analyses. The ROIs are bilateral IFGoper, anterior IPL (aIPL), posterior STG (pSTG), and posterior IPL (pIPL). Specifically, the bilateral activation clusters in IFGoper and aIPL in the conjunction analyses were used to define ROIs of hypothetical motor (simulation) and somatosensory regions operating in the internal forward model (see Fig. 1D). The bilateral activation clusters in pSTG in contrast of OA > SA were used to define auditory ROI because of the auditory speech feedback in the OA task but not in the SA task (see Fig. 1A). Furthermore, the observation of bilateral activation clusters in pIPL (most overlapped with AG) in the contrast of IA > OA may indicate additional processes in internal forward models and hence was defined as another ROI for exploration (see Fig. 1B). Spherical ROIs (6 mm radius) were last defined centering on the activation of these clusters. The MNI coordinates of each ROI center were shown in supplementary materials.

The FC analysis focused on the temporal correlations of time course between each pair of predefined ROIs. In each ROI, the BOLD time courses were first extracted from the preprocessed images and averaged across all the voxels within that ROI. The time courses from gray matter volume were also extracted and the values of all voxels were averaged. The partial correlation coefficient values (Pearson's) was calculated between the time courses in each pair of ROIs while controlling for the time courses of total gray matter for each participant (Salvador et al., 2005). This method provides unique correlation coefficient values between pairs of ROIs with contributions of other areas and autocorrelation canceled out (Fair et al., 2010). The resulting values were further transformed using Fisher's r-z to improve normality.

At the group level, one-sample t-tests were first used to examine significant connectivity (higher than 0) in each task. More importantly, paired samples t-tests were used to examine the possible differences in the resulting correlation values between each pair of tasks. The false discovery rate (FDR) method ($p = 0.05$) was used for the correction of multiple comparisons (Benjamini & Yekutieli, 2001). We illustrated the significant connectivity with both the matrix model and the ball-and-stick model using the BrainNet Viewer software in Figs. 2 and 3 (Xia, Wang, & He, 2013). Furthermore, two-way repeated measures ANOVA was used to examine the potential lateralization of FC between the parietal and pSTG ROIs in internal forward models. For each parietal ROI (aIPL and pIPL), the ANOVA was performed with two factors: laterality of parietal ROI (left and right) and laterality of pSTG (left and right).

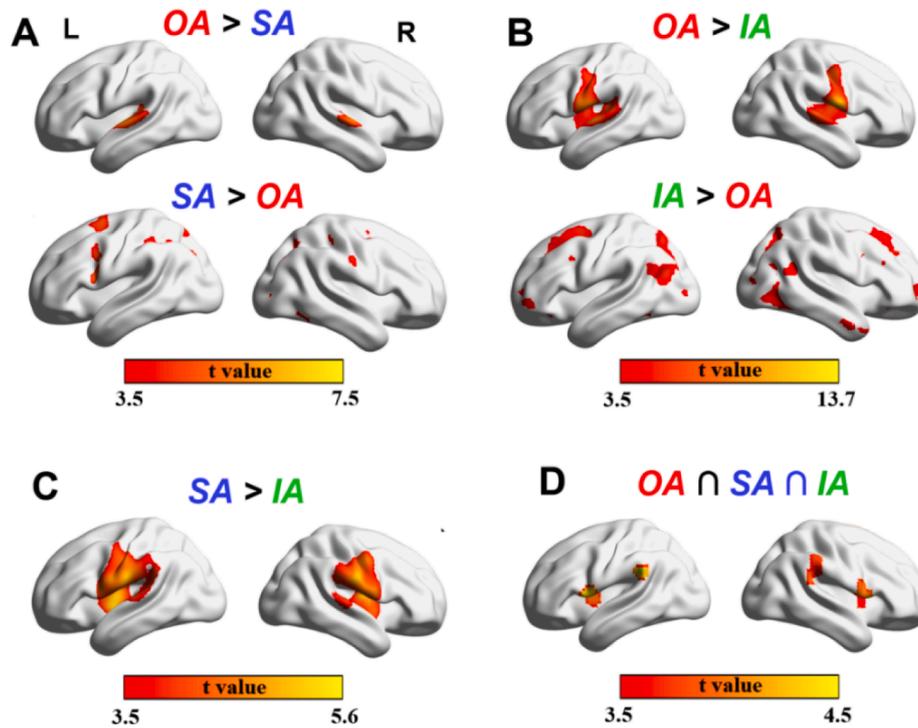


Fig. 1. Activation results of univariate analyses. A-C, the significant activation differences in each pair comparisons of tasks. D, the activation results of conjunction analyses across all three tasks. See supplementary materials for the full list of detailed activation results. OA, overt articulation; SA, silent articulation; IA, imagined articulation. L, left hemisphere; R, right hemisphere.

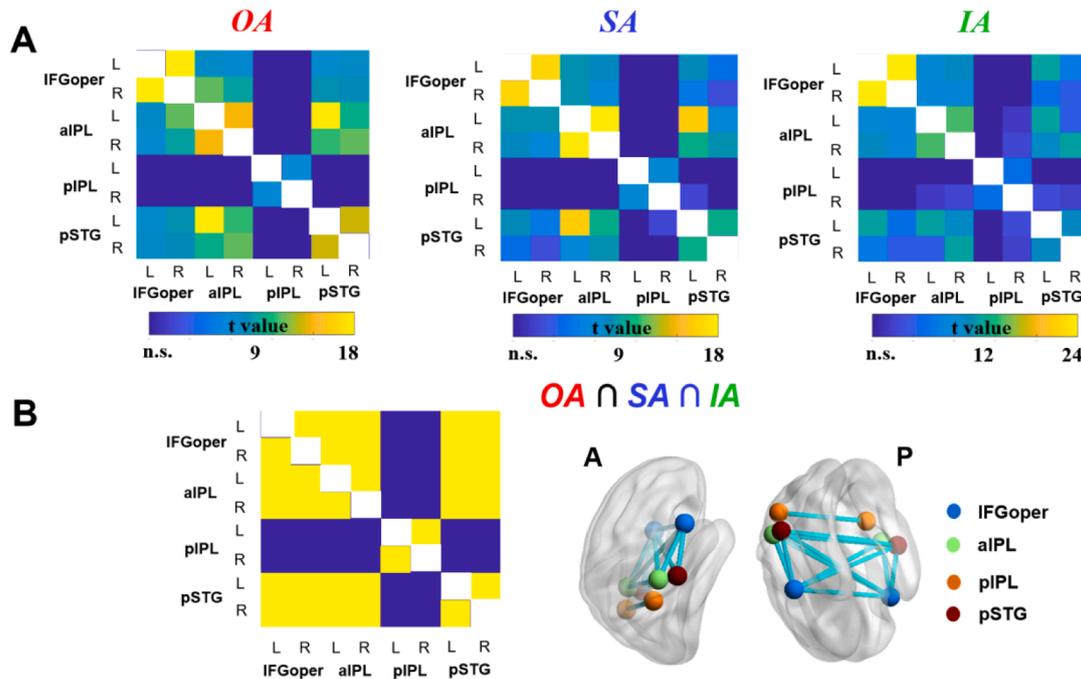


Fig. 2. Significant FC in each task (A) and among all three tasks (B). A, each element in the matrix represents the values between each pair of ROIs. Dark blue denotes the non-significant results in ROI pairs. B, the common significant FC among all three tasks was presented with both matrix and ball-and-stick illustration. In the matrix, yellow denotes significant FC and dark blue denotes the non-significant FC. OA, overt articulation; SA, silent articulation; IA, imagined articulation. A, anterior; P, posterior. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Results

3.1. Univariate results

The detailed activation results of univariate analyses were presented

in supplementary materials. Here, we mainly reported the results that were most relevant to the ROI definition for FC analyses. The OA > SA contrast revealed greater activation in the bilateral STG (see Fig. 1A). The IA > OA contrast showed greater activation in bilateral piPL (Fig. 1B). Moreover, the conjunction analyses across all three tasks

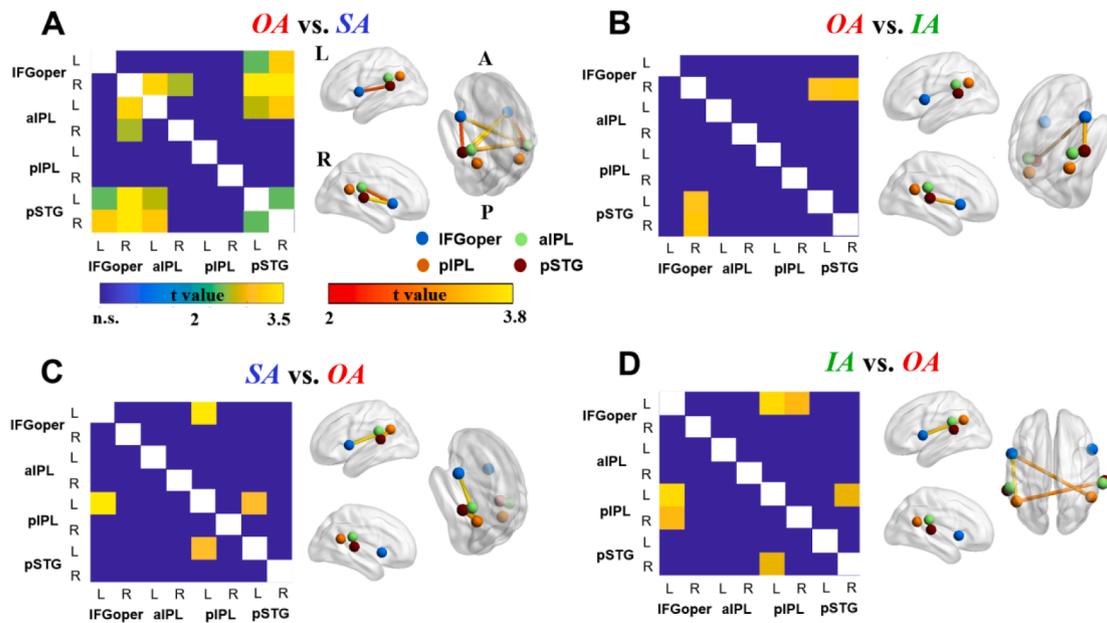


Fig. 3. The pair-comparison results of the task modulation effects on FC in both matrix and ball-and-stick illustration. Dark blue denotes the non-significant results in the compared pair of ROIs. No significant results were found in the contrasts of $SA > IA$ or $IA > SA$. OA, overt articulation; SA, silent articulation; IA, imagined articulation. L, left hemisphere; R, right hemisphere; A, anterior; P, posterior. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

showed bilateral activation in IFGoper and aIPL (Fig. 1D).

3.2. FC results

We first calculated the partial correlation values of FC in each task. As shown in Fig. 2A, the results showed that the FC of one ROI and its contralateral ROI (e.g. left and right IFGoper) was generally stronger than that of other pairs of ROIs. Moreover, in each task, we examined which FC was significantly higher than zero. The results showed that all the FC (except some connectivity with pIPL) were significant in all three tasks. For connectivity with pIPL, only the FC between left and right pIPL was significant in the OA task. In the SA task, the FC between right pIPL and left pSTG was additionally significant. In the IA task, the FC between right pIPL and bilateral aIPL and pSTG were additionally significant. Moreover, we calculated the common significant FC among all three tasks. As shown in Fig. 2B, the FC from IFGoper to aIPL, the FC from IFGoper to pSTG, and the FC from aIPL to pSTG were significant among all tasks. Moreover, the bilateral connectivity of all four ROIs was also significant among all tasks.

We further examined how the strength of FC was modulated by tasks. Paired samples t-tests were first used to examine the FC differences between tasks (see Fig. 3 and Table 1). For the directional contrasts between OA and tasks without auditory feedback (SA and IA), the FC between IFGoper and pSTG in each hemisphere was significantly stronger in OA than those in SA (see Fig. 3A). Similar results of stronger FCs between the right IFGoper and bilateral pSTG were observed in OA compared with those in IA (see Fig. 3B). Moreover, the FCs between bilateral aIPL and right IFG in OA were significantly stronger than those in SA (see Fig. 3A). More importantly, the FC between left aIPL and bilateral pSTG were significantly stronger in OA than those in SA (see Fig. 3A).

Whereas, for the directional contrasts between tasks without auditory feedback (SA and IA) and the task with auditory feedback (OA), the FCs between pIPL and IFGoper and between pIPL and pSTG were significantly stronger in the contrasts of $SA > OA$ and $IA > OA$ (see Fig. 3C and 3D). No significant results were found in the contrasts of $SA > IA$ or $IA > SA$.

The ANOVA results showed different FC patterns of pSTG with aIPL

Table 1

The task modulation effects on FC between ROIs revealed by paired t-tests.

Task Comparison	ROI pairs		t values	Cohen's d
	ROI_1	ROI_2		
OA > SA	LIFG_oper	LpSTG	2.49 *	0.32
	LIFG_oper	RpSTG	3.37 **	0.62
	RIFG_oper	LpSTG	3.46 **	0.69
	RIFG_oper	RpSTG	2.71 *	0.89
	RIFG_oper	LalIPL	8.60 ***	0.48
	RIFG_oper	RalIPL	5.34 ***	0.34
	RIFG_oper	LpSTG	2.77 *	0.45
OA > IA	RIFG_oper	RpSTG	3.26 *	0.58
	RIFG_oper	RpSTG	3.28 *	0.70
	LIFG_oper	LpIPL	3.82 *	0.65
SA > OA	LpIPL	LpSTG	3.17 *	0.31
	LIFG_oper	LpIPL	3.57 *	0.66
IA > OA	LIFG_oper	RpIPL	3.03 *	0.55
	LpIPL	RpSTG	2.92 *	0.53

Note. The connectivity involved ROIs in the parietal lobe (aIPL and pIPL) are highlighted. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$.

and pIPL (see Fig. 4). For aIPL, the results showed a significant main effect of laterality of aIPL ($F(1, 65) = 4.63, p < 0.05, \eta^2 = 0.07$), a significant main effect of laterality of pSTG ($F(1, 65) = 96.28, p < 0.001, \eta^2 = 0.60$), and significant interaction between laterality of aIPL and laterality of pSTG ($F(1, 65) = 215.46, p < 0.001, \eta^2 = 0.77$). For both left and right aIPL, follow-up comparisons showed that the FC with left pSTG was significantly stronger than the FC with right pSTG ($ps < 0.005$), suggesting the left-lateralized speech network.

For pIPL, the results showed a significant main effect of laterality of pIPL ($F(1, 65) = 70.32, p < 0.001, \eta^2 = 0.52$) and significant interaction between laterality of pIPL and laterality of pSTG ($F(1, 65) = 9.31, p < 0.005, \eta^2 = 0.13$). For both left and right pSTG, follow-up comparisons showed that the FC with right pIPL was significantly stronger than the FC with left pIPL ($ps < 0.001$).

The major results of FC analysis are summarized as the following four points. (1) the FC from IFGoper to aIPL, the FC from IFGoper to pSTG,

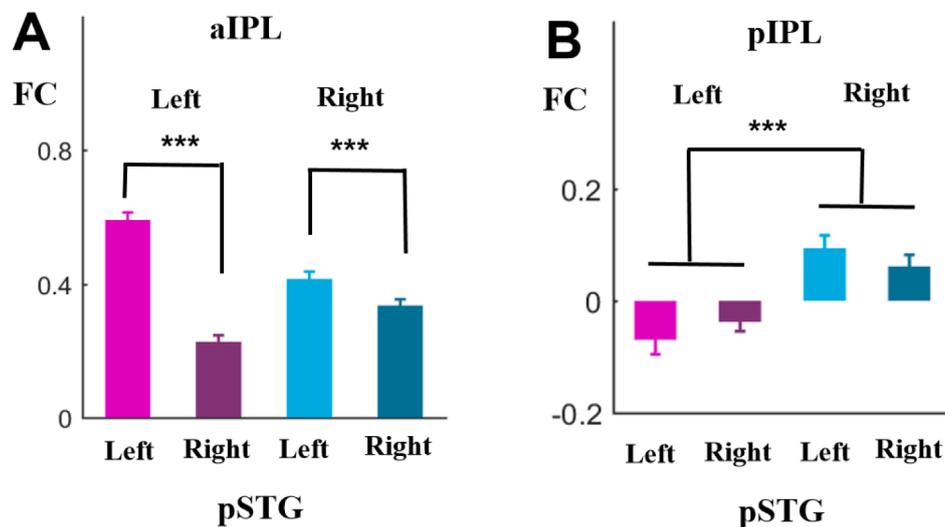


Fig. 4. The intensity of FC between ROIs of parietal regions and pSTG. A, the FC in each pair of ROIs between aIPL and pSTG. B, the FC in each pair of ROIs between pIPL and pSTG. OA, overt articulation; SA, silent articulation; IA, imagined articulation. $***p < 0.005$. Note that significant simple effects of aIPL regard to laterality of pSTG and significant main effects of pIPL laterality are indicated by asterisks. Please refer to the text for all the significant effects.

and the FC from aIPL to pSTG were significant among all tasks. (2) the FC between left aIPL and bilateral pSTG were significantly stronger in the contrast of OA > SA. (3) the FC between pIPL and IFGoper/pSTG were significantly stronger in the contrasts of SA > OA and IA > OA. (4) the FC between the aIPL and left pSTG was stronger than that to the right pSTG.

4. Discussion

The study combined univariate analyses with FC analyses to systematically examine how the somatosensory and auditory estimations in the internal forward models were modulated by speech tasks. The conjunction across overt, silent, and imagined speech tasks showed bilateral activation in IFGoper and aIPL, suggesting that motor simulation and somatosensory estimation operate in the internal forward model during speech production. More importantly, the FC results showed stronger connectivity between aIPL and STG in the OA task compared with the SA task. These results suggest the relationship between somatosensory estimation and auditory estimation can be modulated by speech tasks. Moreover, stronger connectivity between IFGoper and pIPL, and between pIPL and STG were observed in SA and IA tasks compared with the OA task. These results are consistent with the cascaded processing account during speech production.

The study provides evidence for the internal forward model in all speech tasks at both regional and network levels. First, the conjunction analyses across three tasks showed bilateral activation in IFGoper and aIPL. Previous studies suggest that IFGoper played a role in motor preparation and motor simulation (Okada, Matchin, & Hickok, 2018; Park, Thut, & Gross, 2020; Tian et al., 2016), and aIPL is related to sensorimotor and somatosensory estimation (Golfinoopoulos et al., 2011; Rogalsky et al., 2015). Therefore, in overt, silent, and imagined speech tasks, the motor simulation is carried out and a copy of the planned motor commands is sent to the somatosensory areas for estimation. Second, the FC results showed significant connectivity among IFGoper, aIPL, and pSTG in all three tasks, even in the IA task where the movement, somatosensory and auditory feedback are all absent. These results further suggest the existence of a cascaded process among frontal, parietal, and temporal regions that mediates the internal forward models for collaboratively achieving efficient and accurate speech production.

The task-dependent modulation of connectivity reveals the characteristics of auditory and somatosensory estimations, as well as the relation between motor simulation and sensory estimations in the

internal forward models. First, the connectivity between IFGoper and pSTG was stronger in OA than in SA task (Fig. 3, Table 1). In the OA task, participants can hear the external auditory feedback and compare it with the auditory estimation generated from the internal forward models. Therefore, the IFGoper and pSTG need to be more connected for error detection and correction in the OA task. Second, the results showed stronger connectivity between IFGoper and aIPL in OA than those in SA. These results suggest the weights between the motor and somatosensory regions can be adjusted by task demands and the availability of feedback in different modalities.

The modulation of connectivity to temporal auditory regions is further consistent with the existence of a cascaded process in the internal forward model. The stronger connectivity between aIPL and STG in the OA task than in the SA task (Fig. 4A) suggests that the somatosensory estimation and auditory estimation are closely and dynamically related. That is, the “functional connectivity” between the somatosensory and auditory estimations would achieve more collaborative and efficient estimations in the internal forward models. These results of connectivity between aIPL and STG are consistent with findings observed in MEG studies (Li, Luo, & Tian, 2020) and fMRI studies using speech imagery tasks (Chu et al., 2022).

The FC between aIPL and STG in OA was greater than that in SA but was not significantly different from that in IA. This may be because the presence of somatosensory feedback but the absence of auditory feedback in SA promotes the internal forward model in the somatosensory domain in the aIPL but “de-emphasizes” the linking between somatosensory and auditory domains, yielding the strength of the connection between aIPL and STG weaker in SA than that in OA. Whereas, the absence of feedback in both auditory and somatosensory domains desensitizes the “de-emphasis” of the connection between aIPL and STG, which makes the effect comparable in the contrast between OA and IA.

The IPL can be a hub that functionally links the motor and sensory information. The location of IPL is at the junction between the occipital, temporal, and parietal lobes. This region is considered an important interface that conveys and integrates information between different modalities and processing subsystems (Seghier, 2013). Therefore, the relay hub of aIPL in the cascade account is consistent with the proposed role of this parietal region as an interface in the brain network. That is, the efference copy of a motor command can be first sent from frontal motor regions to parietal somatosensory regions for somatosensory estimation. This estimation may be followed by an auditory efference copy that is sent from parietal somatosensory regions to temporal

auditory regions for auditory estimation (Chu et al., 2022; Hickok, 2012; Poeppel, 2014; Tian et al., 2016).

The cascaded process between parietal and temporal lobes provides more specific neural foundations for somatosensory and auditory estimations during regulating speech production. That is, these two estimations work collaboratively and complement each other when both somatosensory and auditory feedbacks are available. This is evident in the behavioral compensation in the auditory feedback perturbation studies. Partial compensation was always observed when perturbing auditory feedback – the adjustment in vocalization was less than the amount of perturbation in the auditory feedback (Chen, Liu, Xu, & Larson, 2007; Liu & Larson, 2007). This could be due to the cascaded process between somatosensory and auditory estimations in internal forward models implemented by the available functional connectivity between parietal and temporal regions. When auditory feedback is on-line altered but somatosensory feedback is intact, because both somatosensory and auditory estimations are available during speech production, conflicts in different modalities are identified by the connectivity between parietal somatosensory and temporal auditory regions. The conflicts limit the compensation in the auditory domain and yield partial behavioral compensation. The connectivity between somatosensory and auditory estimations is also consistent with the negatively correlated compensations in auditory and somatosensory modalities when perturbations were available in both domains (Lametti et al., 2012). That is, the errors in the somatosensory domain can be transmitted to the auditory domain and are taken into consideration when performing the compensation to the perturbed auditory feedback.

Our results are consistent with a cascaded process in the internal forward models during speech production. It needs to be noted that the existence of the cascaded process does not preclude the parallel processes. The parallel and cascaded processes could co-exist and may mediate distinct behaviors and different functions that may operate in various contexts. For example, during development, the causal relations between speech articulation and auditory consequences must be established. That is, the sensorimotor aspects of speech production must link with the auditory aspects. However, the motor status and awareness of articulators cannot be directly accessed but via somatosensory and proprioception (Desmurget et al., 2009; Desmurget & Sirigu, 2009; Haggard, Clark, & Kalogerias, 2002). This indirect access to motor awareness and status necessities the parietal somatosensory processes as an interface to link motor action with its auditory consequences. When the relation between motor and auditory is established after repetitive practice, the direct pathway from the motor system to auditory estimation can be operated independently, paralleling the motor-to-somatosensory estimation pathway. However, in special situations, such as lack of or noisy auditory feedback, the cascaded process would be apparent and influence the processes in the auditory domain and the following speech production control.

Interestingly, contrasting with the differences in connectivity involving aIPL, the activations and connectivity were increased for pIPL in SA and IA tasks than those in the OA task. More activations were found in pIPL in the IA task than in the OA task. Moreover, the connectivity between pIPL and IFGoper/STG was stronger in the SA and IA tasks than in the OA task (see Table 1). One possible explanation is that pIPL is also related to somatosensory estimation during speech production (Zhang et al., 2020). To be specific, the IA > OA contrast showed greater activation in bilateral pIPL (Fig. 1B), which means less activity in the pIPL in the OA task. This may result in more negative FC (anti-correlation) involving pIPL during OA. That is, pIPL could be inhibited when both auditory and somatosensory feedbacks are available. In most usual situations, somatosensory and auditory domains are redundant and speech production is executed correctly in both domains. Therefore, pIPL may complement aIPL that mediates somatosensory estimation and the more negative FC indicates the redundancy between somatosensory and auditory domains, resulting in more weighing over but partial compensation in the auditory domain. If this speculation is correct, our

results would suggest that pIPL may play a compensatory role in somatosensory estimation.

Another possible explanation is that more activations and stronger connectivity in pIPL are related to the role of pIPL in the default mode network (DMN). Previous studies have shown that pIPL (most overlapped with AG) is part of the DMN (Seghier, 2013), which designates a set of cortical regions that are strongly deactivated during goal-directed tasks as compared with rest or passive baselines. In the current study, when considering the behavioral output, SA and IA tasks are closer to rest or passive baseline compared with the OA task. The OA task is more explicit and requires more cognitive resources than the SA and IA tasks. Altogether, our study suggested distinct roles of aIPL and pIPL during speech production. More studies are needed to further examine the potential functional distinctions between aIPL and pIPL.

Our results also suggest the potential lateralization of internal forward models during speech production. Among the FCs in the network of IFGoper, IPL, and pSTG, the majority of the connectivity show left lateralization (see Figs. 3 and 4, Table 1). The temporal auditory estimation regions have stronger connectivity in the left hemisphere. For example, for both left and right aIPL, the FC with left pSTG was significantly stronger than the FC with right pSTG. These results indicate that the internal forward models during speech production probably operated in a left-lateralized manner.

It is worth noting that there are two considerations in the current study. First, although the FC analyses were widely adopted in the literature (Kilteni & Ehrsson, 2020; Sun, Li, Ding, Wang, & Li, 2019), they did not provide directionality between brain regions. Therefore, the directionality of information transmission should be considered with caution in the current study. Although the present results are consistent with this cascaded process, future studies are needed to directly test the directionality of causal influence among brain regions (e.g. using Dynamic Causal Modelling), which can provide stronger evidence for the cascaded process. Secondly, the IA task is a practical and informative protocol to directly examine the neural mechanism underlying internal forward models and widely used in previous studies (Neef et al., 2016; Tian et al., 2018). However, it is difficult to empirically check the outcome of the IA task because of the lack of behavioral measures in mental imagery tasks, especially in the auditory domain. Therefore, it is hard if not impossible to behaviorally verify whether participants imagined articulating each syllable as required. In the current study, the univariate and FC analyses of the IA task showed significant activations and connectivity reflecting imagined articulating (see Figs. 1 and 2) that are consistent with previous studies. Therefore, the objective neural results support that participants conducted the IA task as required.

In conclusion, using FC analysis with different speech tasks, we found a common neural network among frontal motor, parietal somatosensory, and temporal auditory regions that mediate internal forward models during speech production. The connectivity strengths in the internal forward network, especially the ones with hypothetical somatosensory and auditory estimations in parietal and temporal lobes, are modulated by task demand. These results support the task-dependant communication among related brain regions in the internal forward model, which is consistent with a cascaded process during speech production.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Qian Chu for his help with data analyses.

Funding

This study was supported by the National Natural Science Foundation of China 32071099 and 32271101, Natural Science Foundation of Shanghai 20ZR1472100, Program of Introducing Talents of Discipline to Universities, Base B16018, NYU Shanghai Boost Fund to XT, Natural Science Foundation in Shaanxi Province of China 2023-JC-QN-0232, and Social Science Foundation in Shaanxi Province of China 2022K015 to WZ.

References

- Behroozmand, R., & Larson, C. R. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, *12*(1), 54.
- Behroozmand, R., Liu, H., & Larson, C. R. (2011). Time-dependent neural processing of auditory feedback during voice pitch error detection. *Journal of Cognitive Neuroscience*, *23*(5), 1205–1217.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, *29*(4), 1165–1188.
- Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (1998). Central cancellation of self-produced tickle sensation. *Nature Neuroscience*, *1*(7), 635–640.
- Chen, S. H., Liu, H., Xu, Y., & Larson, C. R. (2007). Voice F0 responses to pitch-shifted voice feedback during English speech. *Journal of Acoustic Society of America*, *121*(2), 1157–1163.
- Chu, Q., Ma, O., Hang, Y., & Tian, X. (2022). Dual-stream cortical feedbacks mediate sensory prediction. *bioRxiv*.
- Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science*, *324*(5928), 811–813.
- Desmurget, M., & Sirigu, A. (2009). A parietal-premotor network for movement intention and motor awareness. *Trends in Cognitive Sciences*, *13*(10), 411–419.
- Duffau, H., Gatignol, P., Denvil, D., Lopes, M., & Capelle, L. (2003). The articulatory loop: Study of the subcortical connectivity by electrostimulation. *NeuroReport*, *14*(15).
- Fair, D., Bathula, D., Mills, K., Costa Dias, T., Blythe, M., Zhang, D., ... Nagel, B. (2010). Maturing thalamocortical functional connectivity across development. *4*(10).
- Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *Neuroimage*, *55*(3), 1324–1338.
- Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*(3), 594.
- Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280–301.
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382–385.
- Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145.
- Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407–422.
- Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology*, *33*, 174–181.
- Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, *5*, 82.
- Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience*, *14*(8), 1125–1138.
- Kiltner, K., & Ehrsson, H. H. (2020). Functional connectivity between the cerebellum and somatosensory areas implements the attenuation of self-generated touch. *Journal of Neuroscience*, *40*(4), 894–906.
- Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *The Journal of Neuroscience*, *32*(27), 9351.
- Li, Y., Luo, H., & Tian, X. (2020). Mental operations in rhythm: Motor-to-sensory transformation mediates imagined singing. *PLoS Biology*, *18*(10), e3000504.
- Liu, H., & Larson, C. R. (2007). Effects of perturbation magnitude and voice F0 level on the pitch-shift reflex. *Journal of Acoustical Society of America*, *122*(6), 3671–3677.
- Neef, N. E., Büttner, C., Anwender, A., Friederici, A. D., Paulus, W., & Sommer, M. (2016). Left posterior-dorsal area 44 couples with parietal areas to promote speech fluency, while right area 44 activity promotes the stopping of motor responses. *Neuroimage*, *142*, 628–644.
- Okada, K., Matchin, W., & Hickok, G. (2017). Neural evidence for predictive coding in auditory cortex during speech production. *Psychonomic Bulletin & Review*, *25*(1), 423–430.
- Okada, K., Matchin, W., & Hickok, G. (2018). Phonological feature repetition suppression in the left inferior frontal gyrus. *Journal of Cognitive Neurosciences*, *30*(10), 1549–1557.
- Park, H., Thut, G., & Gross, J. (2020). Predictive entrainment of natural speech through two fronto-motor top-down channels. *Language, Cognition Neuroscience*, *35*(6), 739–751.
- Poeppl, D. (2014). The neuroanatomic and neurophysiological infrastructure for speech and language. *Current Opinion in Neurobiology*, *28*, 142–149.
- Rogalsky, C., Poppa, T., Chen, K. H., Anderson, S. W., Damasio, H., Love, T., & Hickok, G. (2015). Speech repetition as a window on the neurobiology of auditory-motor integration for speech: A voxel-based lesion symptom mapping study. *Neuropsychologia*, *71*(01), 18–27.
- Salvador, R., Suckling, J., Coleman, M. R., Pickard, J. D., Menon, D., & Bullmore, E. J. C. (2005). Neurophysiological architecture of functional magnetic resonance images of human brain. *15*(9), 1332–1342.
- Seghier, M. L. (2013). The angular gyrus: Multiple functions and multiple subdivisions. *Neuroscientist A Review Journal Bringing Neurobiology Neurology & Psychiatry*, *19*(1), 43–61.
- Smith, D. J., Stepp, C., Guenther, F. H., & Kearney, E. (2020). Contributions of auditory and somatosensory feedback to vocal motor control. *Journal of Speech, Language, and Hearing Research*, *63*(7), 2039–2053.
- Sun, X., Li, L., Ding, G., Wang, R., & Li, P. (2019). Effects of language proficiency on cognitive control: Evidence from resting-state functional connectivity. *Neuropsychologia*, *129*, 263–275.
- Tian, X., Ding, N., Teng, X., Bai, F., & Poeppel, D. (2018). Imagined speech influences perceived loudness of sound. *Nature Human Behaviour*, *2*(3), 225–234.
- Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, *1*(3), 255–262.
- Tian, X., & Poeppel, D. (2012). Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, *6*(6), 314.
- Tian, X., & Poeppel, D. (2013). The effect of imagination on stimulation: the functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience*, *25*(7), 1020–1036.
- Tian, X., & Poeppel, D. (2015). Dynamics of self-monitoring and error detection in speech production: Evidence from mental imagery and MEG. *Journal of Cognitive Neuroscience*, *27*(2), 352–364.
- Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*, *77*(7), 1–12.
- Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage*, *39*(3), 1429–1443.
- Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, *3 Suppl*(Suppl), 1212–1217.
- Xia, M., Wang, J., & He, Y. (2013). BrainNet viewer: A network visualization tool for human brain connectomics. *PLoS ONE*, *8*(7), e68910.
- Zhang, W., Liu, Y., Wang, X., & Tian, X. (2020). The dynamic and task-dependent representational transformation between the motor and sensory systems during speech production. *Cognitive Neuroscience*, 1–11.