

Research



Cite this article: Valeriani D, Simonyan K. 2021 The dynamic connectome of speech control. *Phil. Trans. R. Soc. B* **376**: 20200256. <https://doi.org/10.1098/rstb.2020.0256>

Accepted: 10 June 2021

One contribution of 21 to a theme issue ‘Vocal learning in animals and humans’.

Subject Areas:

neuroscience

Keywords:

speech, brain networks, effective connectivity

Author for correspondence:

Kristina Simonyan

e-mail: kristina_simonyan@meei.harvard.edu

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5527040>.

The dynamic connectome of speech control

Davide Valeriani^{1,2} and Kristina Simonyan^{1,2,3}

¹Department of Otolaryngology – Head and Neck Surgery, Massachusetts Eye and Ear, 243 Charles Street, Boston, MA 02114, USA

²Department of Otolaryngology-Head and Neck Surgery, Harvard Medical School, 243 Charles Street, Boston, MA 02114, USA

³Department of Neurology, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA

DV, 0000-0001-9866-0063; KS, 0000-0001-7444-0437

Speech production relies on the orchestrated control of multiple brain regions. The specific, directional influences within these networks remain poorly understood. We used regression dynamic causal modelling to infer the whole-brain directed (effective) connectivity from functional magnetic resonance imaging data of 36 healthy individuals during the production of meaningful English sentences and meaningless syllables. We identified that the two dynamic connectomes have distinct architectures that are dependent on the complexity of task production. The speech was regulated by a dynamic neural network, the most influential nodes of which were centred around superior and inferior parietal areas and influenced the whole-brain network activity via long-ranging coupling with primary sensorimotor, prefrontal, temporal and insular regions. By contrast, syllable production was controlled by a more compressed, cost-efficient network structure, involving sensorimotor cortico-subcortical integration via superior parietal and cerebellar network hubs. These data demonstrate the mechanisms by which the neural network reorganizes the connectivity of its influential regions, from supporting the fundamental aspects of simple syllabic vocal motor output to multimodal information processing of speech motor output.

This article is part of the theme issue ‘Vocal learning in animals and humans’.

1. Introduction

Speech production is a uniquely complex human behaviour that requires numerous brain regions to perceive, process and comprehend the sensory input, integrate it with cognitive and motor intent, and execute the synchronized movement of over 100 orofacial, laryngeal and respiratory muscles. Owing to its complexity, neuroimaging studies of speech control have primarily focused on examining distinct components of the speech network, such as motor output [1–4], verbal fluency [5,6], phonological processing [7,8] or sensorimotor integration [9–11]. A few studies have investigated functional relationships between regions at the whole-brain level as well as interactions between different speech network components, describing the large-scale functional connectome of speech control [12–14]. Compared with other, more simplistic motor and non-motor behaviours, speech production was shown to require a specialized network with preferential recruitment of prefrontal, inferior parietal and cerebellar regions. However, while these studies mapped the complexity of global speech network organization, the directionality of regional interactions and thus information transfer within the speech connectome remain unknown.

Rooted in methodological limitations, examinations of directional (or dynamic, effective) network connectivity have been typically limited to assessing the interactions only between a few (4 to 6) brain regions owing to an unfeasibly high computationally cost of modelling the whole-brain network [15]. Recently, analytical advances in network neuroscience have made possible the

development of a computationally efficient approach of regression dynamic causal modelling (rDCM), which interrogates effective connectivity of the large-scale brain network, including over 200 regions [16]. Leveraging this methodology, we employed rDCM together with graph-theoretical analysis to examine the whole-brain directional connectomes during the production of grammatically correct English sentences, as examples of meaningful real-life speech, and production of syllables, as examples of meaningless learned motor vocal output. Our experimental design was similar to our prior study [12] in order to allow a comparative assessment between functional and effective connectomes controlling syllable and speech production. Our overarching hypothesis was that the speech production network (SPN) exhibits a more complex integration of cortical sensorimotor information transfer, directly influencing the activity of the primary motor cortex for the output of meaningful speech, whereas the syllable production network (SylPN) is characterized by directional coupling between cortical and subcortical areas to support more fundamental aspects of voluntary vocal motor output.

2. Material and methods

(a) Study participants

We recruited 36 monolingual native-English-speaking healthy individuals (22 females/14 males; age 51.4 ± 10.6 years). All participants were right handed as determined by the Edinburgh Handedness Inventory and had no history of any past or present neurological, psychiatric, otolaryngological, or developmental speech and language problems. Data from 14 participants were used in the previous study of functional connectome of speech [12], whereas the remaining 22 participants were recruited specifically for this study. All participants had a normal cognitive function and scored at least 27 points on the Mini-Mental State Examination. All participants provided written informed consent, which was approved by the Institutional Review Board of Icahn School of Medicine at Mount Sinai and Mass General Brigham.

(b) Experimental design and MRI data acquisition

Brain images for all subjects were acquired on a 3.0 Tesla Philips MRI scanner equipped with an eight-channel head coil. At the beginning of the experimental session, a high-resolution T1-weighted whole-brain image was collected for anatomical reference using a three-dimensional magnetization-prepared rapid acquisition gradient echo sequence (TR = 4.5 ms, TE = 3.4 ms, flip angle = 8°, 172 slices, slice thickness = 2 mm). Whole-brain functional images were obtained using gradient-weighted echo-planar imaging (EPI) pulse sequences (effective TR = 10.6 s, with 8.6 s for task and 2 s for image acquisition, TE = 30 ms, flip angle = 90°, FOV = 240 mm, voxel size = 3.75×3.75 mm, 36 slices, slice thickness = 4 mm) with blood oxygen level-dependent (BOLD) contrast and event-related sparse-sampling design to minimize motion artefacts. Subjects were instructed to remain motionless during the scanning and to minimize orofacial movements during speaking; their head was tightly cushioned within the head coil to prevent motion artefacts.

The fMRI design included syllable and sentence production and resting as a baseline condition. Specifically, subjects were instructed to produce a syllable sequence /i/i/ to capture brain activity associated with simple vocal motor behaviour and meaningful, grammatically correct English sentences (e.g., 'My father has a new car', 'He is hiding behind the house', 'Sally fell asleep in a soft chair') to assess brain activity associated

with speaking. Eight different sentences were used to minimize the working memory build-up throughout the scanning session while capturing the spectrum of phonological and lexical elements characterizing real-life speech. Each subject completed four scanning runs within the same fMRI session. Each run included 8 syllable trials, 8 speech trials and 16 resting trials, totalling 32 trials of syllable production, 32 trials of sentence production and 64 trials of resting. All trials were presented in a pseudorandomized order. Subjects first listened through the MRI-compatible headphones to four repetitions of the sample syllable or one sample sentence for 3.6 s and then repeated them for 5 s (figure 1a). No stimuli were presented during the resting condition, when subjects rested with their eyes open.

(c) Data preprocessing

Image preprocessing was performed using AFNI software following a standard analytical protocol. Briefly, after discarding the first two volumes for equilibrium correction, all volumes were registered to the high-resolution anatomical scan, spatially smoothed with a 4 mm Gaussian filter and normalized to the AFNI standard Talairach–Tournoux space. To control for motion artefacts, six motion parameters were estimated during the realignment of the EPI volumes and included as covariates of no interest in the regressor, together with three quadratic polynomials that were used to model baseline drifts for each imaging run. In addition, TRs with the Euclidean norm of the motion derivative of greater than or equal to 1.0 were censored out, and censoring of outlier TRs was performed to ensure the stringent removal of TRs containing residual motion artefacts, as described previously [17].

A general linear model analysis identified brain activity related to speech and syllable production, respectively (SPM12 running on Matlab R2018a). A single regressor per task was modelled using a boxcar function tied to the duration of the production. The auditory input associated with the task sample stimuli preceding syllable/speech production was not modelled given the sparse-sampling event-related fMRI design. The task regressor was convolved with a haemodynamic response function and entered into a multiple regression model to predict the observed BOLD response in each voxel. The individual anatomical image was parcellated into 212 regions of interest (ROIs) based on the cytoarchitectonic maximum probability maps and macrolabel atlas [12,18]. The brain parcellation included 142 cortical, 36 subcortical and 34 cerebellar regions (figure 1b). The average BOLD signal during each task production was extracted from each ROI and used in rDCM analysis.

(d) Effective connectivity analysis

(i) Regression dynamic causal modelling

As one of the major approaches to examining effective connectivity, DCM is a generative modelling framework for estimating latent neural states from the measured neural signal [15]. In its linear form, DCM describes neuronal activity dynamics as the result of the directional (effective) connectivity between neuronal populations:

$$\frac{dx}{dt} = Ax + Cu,$$

where x represents neuronal activity, A is the endogenous connectivity and C indicates how external manipulations directly affect neuronal activity. This is the inverse of the traditional haemodynamic-forward model, which maps neuronal dynamics dx/dt to observed BOLD signal. Inverting the generative model relies on a variational Bayesian approach under the Laplace approximation [19], a computationally expensive problem for large

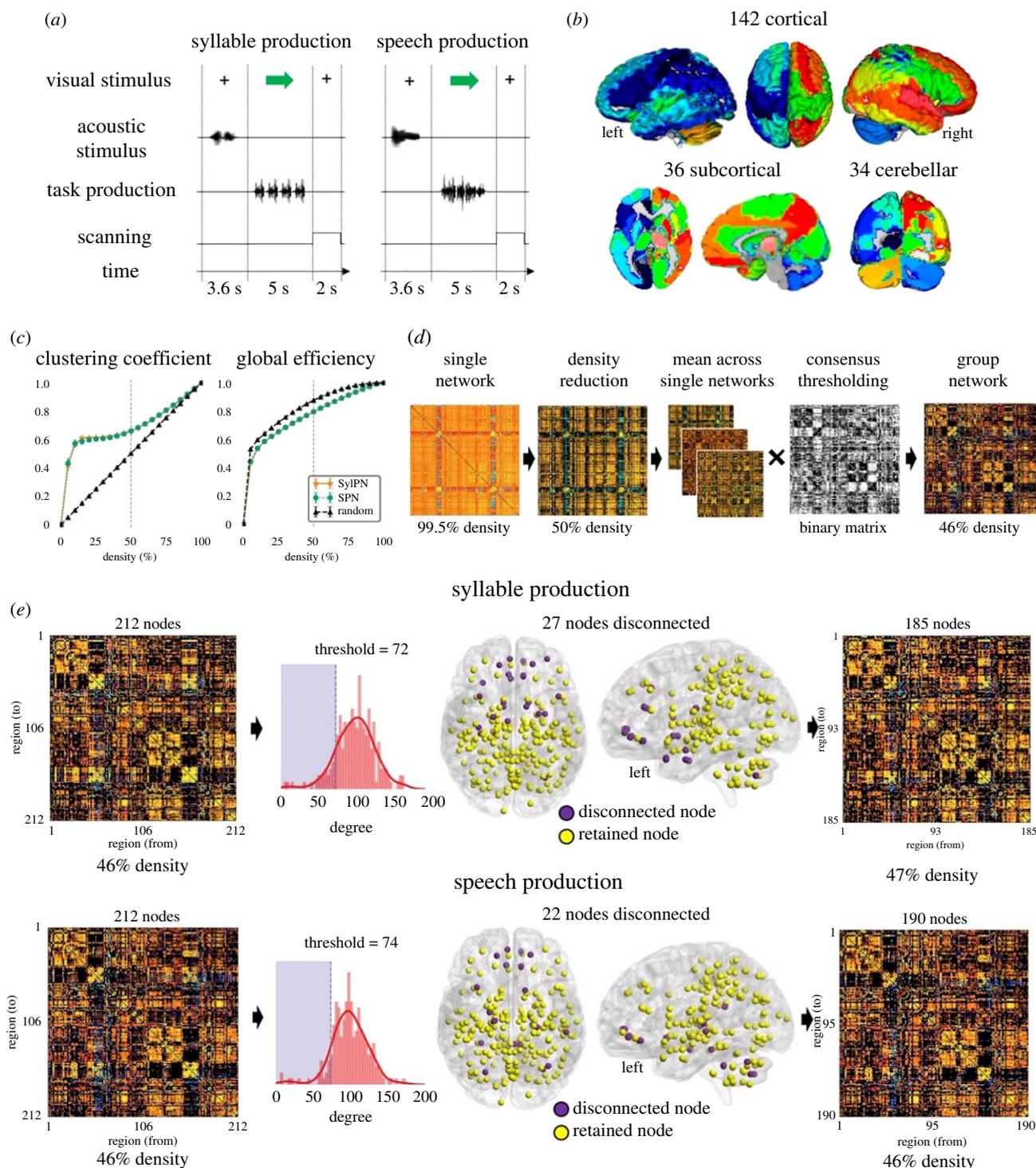


Figure 1. Overview of the experimental design and data processing pipeline. (a) Schematic illustrations of the experimental paradigms for speech and syllable production tasks. (b) Whole-brain parcellations using the cytoarchitectonic maximum probability map and macrolabel atlases composed of 142 cortical, 36 subcortical and 34 cerebellar regions of interest (ROIs). ROIs are visualized using Mango 4.1 software. (c) Clustering coefficient and global efficiency of random and empirical networks at different levels of density. The dashed lines depict the metrics values at 50% density. (d) Schematic of the fMRI data processing pipeline. (e) Schematic of the nodal elimination strategy employed to disconnect low-degree nodes from the group network. The histogram shows the distribution of nodal degrees across the 212 ROIs. Nodes with a degree lower than one standard deviation from the average degree were disconnected (22 nodes, shaded purple area). The final group matrix included 185 nodes for syllable production and 190 nodes for speech production. Disconnected and retained nodes are shown on three-dimensional brain renderings using BrainNet Viewer software.

networks. However, recently introduced rDCM translates state and observation equations from time to frequency domain:

$$\widehat{\frac{dx}{dt}} = A\hat{x} + C\hat{u},$$

where the hat symbol denotes the Fourier transform. This approach makes the likelihood function mathematically tractable and capable of analysing whole-brain networks [16]. Moreover, rDCM assumes that the connectivity parameters are independent

across regions, allowing model inversion to take place one region at a time and, hence, processed in parallel, thus increasing the computational efficiency of rDCM.

(ii) Individual regression dynamic causal modelling network construction and thresholding

To examine the directionality of connections in individual whole-brain effective connectomes during syllable and speech

production, we first explicitly modelled connectivity between all 212 ROIs via reciprocal links (i.e. full A matrix) and assumed that the driving input could elicit activity in all regions (i.e., full C matrix). Empirical validation of rDCM connectivity estimates in a whole-brain network, including more than 200 regions and more than 40 000 parameters, has been shown previously [20]. In each subject, rDCM estimated over 45 000 rDCM parameters within a 212×212 matrix of SPN and SylPN, respectively, using the TAPAS v. 3.1.0 toolbox running on Matlab R2018a. The density of each individual matrix per task was computed as the percentage of non-zero values, which yielded a density of 99.5% for all matrices, i.e., all 212×212 connections except self-connections. The strength of each connection was normalized to the range 0 to 1 in each direction of a reciprocal connection.

Next, because high-density networks tend to exhibit random network characteristics [21] and are difficult to interpret [22], we reduced the density of each individual matrix by absolute value, while keeping only the strongest connections. To determine the density reduction threshold, we constructed 36 synthetic networks (one per subject) for each density starting from 5% to 100%, with 5% incremental steps. A total of 720 synthetic networks (36 subjects \times 20 density levels) were generated. Random connectivity values between 212 nodes of each synthetic network were drawn from a uniform distribution between 0 and 1. With empirical and synthetic SPN and SylPN in the same range of densities, we computed clustering coefficient and global efficiency of real and random networks at different density levels. Clustering coefficient represents how well the node is connected to its neighbours and is computed as the geometric average of the edge weights in the triangles around a node to assess the extent of local community formation [23]. Global efficiency is calculated as the average inverse shortest path length in the network and represents how well nodes are connected to each other. These metrics were averaged across subjects and are reported in figure 1c. The Mann–Whitney rank test was used to compare clustering coefficient and global efficiency between empirical and random networks at Bonferroni-corrected $p < 0.01$. The final individual network density threshold was set to 50% by sorting connections based on their absolute strength with subsequent removal of weakest connections until the density threshold of 50% was reached (figure 1d). This approach allowed reduction of the individual matrix density while ensuring each network exhibited significantly different clustering coefficient and global efficiency compared with random networks. Both empirical networks showed significantly higher clustering coefficient than random networks (SylPN: 0.664 ± 0.025 ; SPN: 0.659 ± 0.028 ; random network: 0.502 ± 0.000 ; $U = 1296$, two-sided $p = 3 \times 10^{-13}$, Mann–Whitney rank test) and significantly lower global efficiency (both SylPN and SPN: 0.794 ± 0.014 ; random network: 0.876 ± 0.001 ; $U = 0$, two-sided $p = 3 \times 10^{-13}$, Mann–Whitney rank test). These findings were similar to differences between empirical and random functional connectivity matrices in our earlier study using Pearson’s correlation coefficients [12], suggesting the overall robustness of SPN and SylPN architectures.

(iii) Group network construction and thresholding

To generate group networks for each task while ensuring the stability of the parameter estimates in our study, we averaged individual thresholded matrices across all subjects and applied distance-dependent consensus thresholding [24] (figure 1d). This thresholding approach creates a consensus network by computing the frequency of each connection across the individual network. It then splits the connections in M bins based on their length (i.e., distance between two nodes) percentiles, where M is set to be the average number of connections in the consensus matrix and selects links with the highest consensus across participants. This procedure generates group networks with

connections present across the majority of participants and approximately the same length distribution of individual networks. In our dataset, the application of distance-dependent consensus thresholding resulted in group networks with densities of 46% ($M = 17$) for each SylPN and SPN. As described previously [12,14], we additionally removed sparsely connected nodes, which had an overall degree (i.e., the total number of nodal connections) below one standard deviation compared with the average nodal degree in each respective group network (figure 1e). This nodal elimination strategy removed 27 nodes from the SylPN and 22 nodes from the SPN and helped reduce the number of weakly connected regions within the network, further decreasing the number of parameters to be estimated in the subsequent analysis. The final group SylPN was composed of 185 nodes at 47% density, and the final group SPN included 190 nodes at 46% density.

(iv) Quantitative analysis of network topology

To examine global features of the effective network architecture, we computed the topological distribution of neural communities (modules), defined as a group of nodes densely connected within their neural community but sparsely connected with nodes in other neural communities [25]. We assessed the optimal community architecture of group SPN and SylPN networks using a heuristic modularity maximization strategy based on the Kernighan–Lin algorithm [26], which employed the Louvain community detection algorithm [27] implemented in the Brain Connectivity Toolbox [28]. To account for the stochastic nature of the heuristic modularity maximization routine, which randomly permuted nodal community assignments, we computed the community structure 1000 times for each network to ensure the robustness of the final modular decomposition [12]. The final community affiliation was determined by quantifying the frequency with which each node was assigned to the same neural community in each iteration. Community analysis was performed separately for inhibitory and excitatory nodes in each network to investigate the topology of inhibitory and excitatory influences in each neural community.

(v) Excitatory and inhibitory nodal influence

To examine excitatory and inhibitory regional influences within SPN and SylPN networks, we computed the inner and outer degree (i.e., the number of incoming and outgoing connections of each node, respectively) and strength (i.e., the weighted sum of incoming and outgoing links of each node, respectively) of each network node. We classified each node as inhibitory or excitatory depending on whether its normalized inner strength was higher or lower than its normalized outer strength [16] in order to highlight the main role played by that node in the network.

Because over 16 000 connections characterized a group network of 190 nodes at 46% density, we limited our analysis of excitatory/inhibitory nodal influences to the strongest t connections within each SylPN and SPN, where t varied between 1 and 10% in steps of 1%. At each iteration, we computed a difference matrix where the connection between two nodes a, b had the following value:

$$w_{\text{diff}}(a,b) = \begin{cases} +1, & \text{if } w_{\text{SPN}}(a,b) \neq 0 \text{ and } w_{\text{SylPN}}(a,b) = 0 \\ -1, & \text{if } w_{\text{SPN}}(a,b) = 0 \text{ and } w_{\text{SylPN}}(a,b) \neq 0 \\ 0, & \text{elsewhere,} \end{cases}$$

where $w_{\text{SPN}}(a,b)$ is the weight of the connection between nodes a and b in the SPN filtered with threshold t , and $w_{\text{SylPN}}(a,b)$ is the corresponding weight in the SylPN. To select the optimal value of t , we computed each node’s degree in the difference matrix and calculated the number of nodes with a positive degree for each threshold value t . We chose the minimum t that allowed us to obtain a difference matrix with at least 25% of the original

212 nodes, capturing the large-scale network specificity while maintaining the differences visualizable. This process set t to 1%. With the chosen t , we then visualized the difference matrix as a connectogram. We computed the degree of specificity for each node as the difference between the number of SPN-specific and SyLPN-specific connections involving that node, regardless of their directionality. For example, if a node was involved in two SPN-specific inhibitory connections and two SyLPN-specific excitatory connections, its degree of specificity would be 0. This measure represented the degree to which each node was involved in network-specific high-strength connections.

(vi) Hub analysis

Network hubs were determined as nodes whose strength and degree were both one standard deviation higher than the average nodal strength and degree in the respective network. The nodal participation coefficient pc_i was computed to measure the distribution of nodal connections among all neural communities in each network. Nodes with $pc_i \geq 20\%$ of maximum participation coefficient of a network with m neural communities ($pc_{\max} = 1-1/m$ ($pc_{\max} = 0.667$ for SyLPN and $pc_{\max} = 0.75$ for SPN)) were classified as connector hubs (i.e. hubs connecting different communities), while nodes with $pc_i < 20\%$ were classified as provincial hubs (i.e. hubs connecting nodes within a community). The similarity of community structures was quantified by estimating the partition distance, p_d , which was calculated as the normalized mutual information between the community affiliation vectors. To assess the extent of each hub's influence within the network depending on the complexity of syllable versus speech production, we examined the length of connectivity distance of each hub in the SyLPN and SPN, respectively, by computing the Euclidean distance of the farthest node connected to each hub.

Finally, to analyse the hub connectivity patterns within the SyLPN and SPN, we considered both shared and distinct hubs in each network. We computed the hub connectivity matrix of the SyLPN and SPN separately by extracting each hub's directional connectivity with other hubs within the network. Inhibitory and excitatory connections within each network were normalized using min-max scaling. High-strength connections were defined as connections with a strength higher than one standard deviation above the average excitatory or inhibitory connection strength in the hub connectivity matrix. If a pair of hubs exhibited both high-strength inhibitory and excitatory connections, we considered the stronger of the two connections.

Network analysis was performed using Python 3.6.0 with the NetworkX 2.4 package [29] and Matlab R2018a with the Brain Connectivity Toolbox [28]. The difference matrix was visualized using Circos software [30].

3. Results

(a) Overall topology of the syllable and speech production networks

The syllable production network (SyLPN) consisted of three neural communities (figure 2a; electronic supplementary material, table S1), predominantly spanning

- (I) bilateral middle and medial frontal gyri, primary motor, premotor and occipital cortex, left superior frontal gyrus, inferior and superior parietal cortex;
- (II) bilateral thalamus (prefrontal, temporal and parietal subdivisions), hippocampus (right area CA, left FD, bilateral area HATA and SUB), amygdala, right inferior frontal cortex, somatosensory cortex, parietal

operculum, auditory cortex, inferior/superior parietal cortex, insula and cerebellum; and

- (III) bilateral basal ganglia, red nucleus, thalamus (premotor, motor, temporal and somatosensory subdivisions), hippocampus (left area CA and right area FD), left inferior frontal, primary somatosensory and auditory cortex, parietal operculum, insula, cerebellum and right superior frontal gyrus.

Out of 185 nodes of the SyLPN, 10% formed the inhibitory subnetwork (i.e., composed of only inhibitory nodes), and 90% of nodes contributed to the excitatory subnetwork (i.e. composed of only excitatory nodes), each consisting of three neural communities (figure 2b,c; electronic supplementary material, tables S2 and S3).

(i) Speech production network

Although the SPN showed moderate similarity in neural community organization compared with the SyLPN ($p_d(\text{SPN}, \text{SyLPN}) = 0.590$), speech production involved overall more complex network architecture and hub distribution. Specifically, the SPN was characterized by four neural communities (figure 2d; electronic supplementary material, table S1), predominantly including

- (I) left inferior/superior frontal and right middle frontal gyri, right somatosensory, inferior/superior parietal and auditory cortex, parietal operculum, insula, hippocampus, medial globus pallidus, cerebellum, bilateral thalamus (premotor, motor, temporal subdivisions) and amygdala;
- (II) bilateral medial and left middle frontal, premotor, primary motor, cingulate and occipital cortex, cuneus, precuneus, caudate nucleus, left inferior/superior parietal cortex, insula, thalamus (parietal subdivision) and cerebellum;
- (III) right inferior/superior frontal gyri, left primary somatosensory and auditory cortex, parietal operculum, bilateral putamen, subthalamic nucleus, thalamus (prefrontal, somatosensory, visual subdivisions), left medial globus pallidus, red nucleus and cerebellum; and
- (IV) bilateral inferior parietal cortex (area hIP2), globus pallidus (lateral), cerebellar lobule VI, right amygdala (area CM) and red nucleus.

Out of 190 nodes in the SPN, the inhibitory subnetwork recruited 10% of nodes, while the remaining 90% of nodes contributed to the excitatory subnetwork, each forming four neural communities (figure 2e,f; electronic supplementary material, tables S2 and S3).

(b) Network-specific connectivity of syllable and speech production networks

More than 25% of nodes in the SyLPN and SPN contributed to the top 1% of strongest connections of each respective network and formed a specialized pattern of the neural organization supporting the production of a given behaviour (figure 3a).

The SyLPN was characterized by distinct connections of the left prefrontal cortex with the right insula, auditory cortex, cerebellum and left nucleus accumbens; right prefrontal cortex

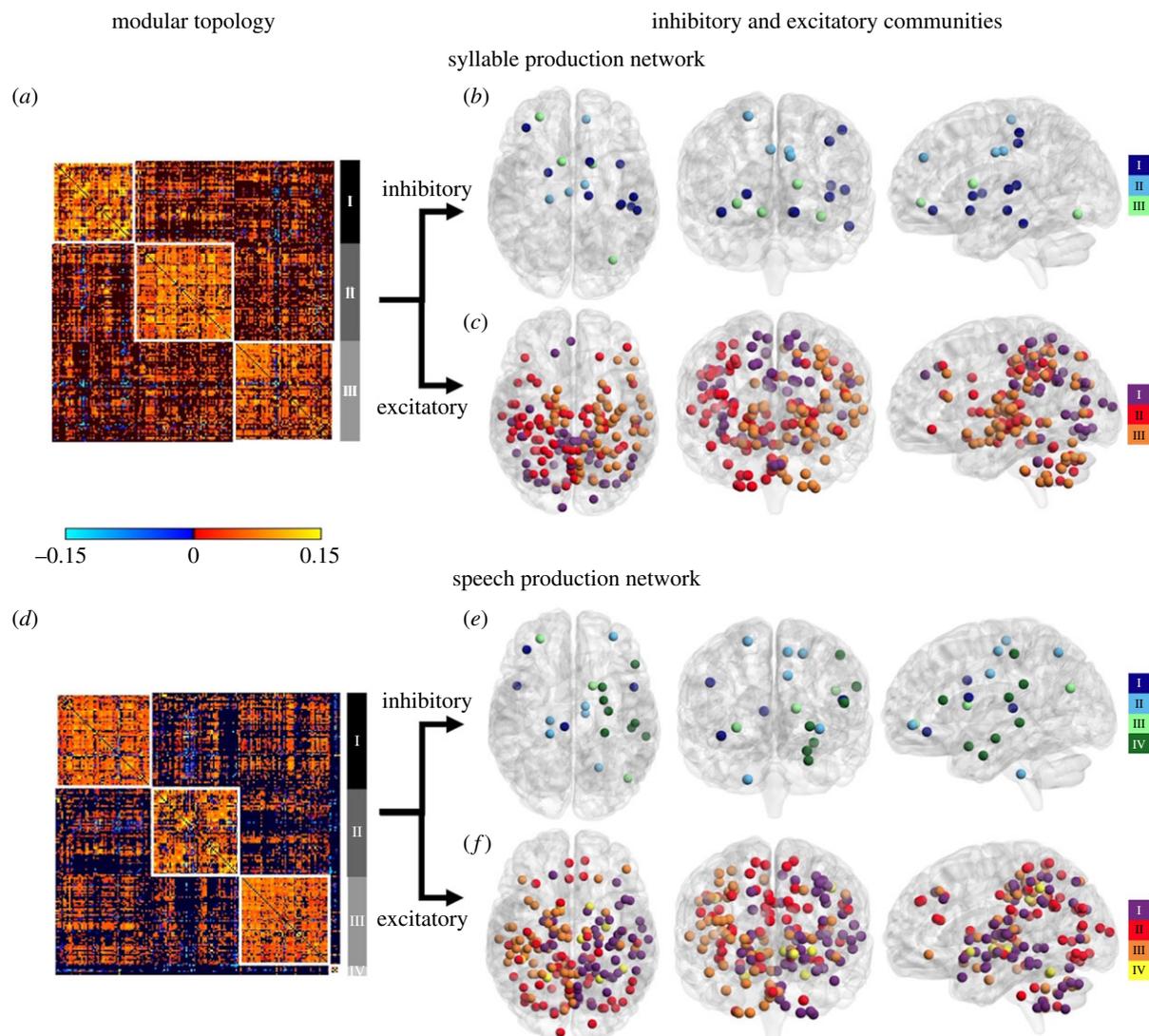


Figure 2. Neural community architecture of the group networks during syllable and speech production. (a,d) Group connectivity matrix of the SyIPN and SPN showing the neural communities identified by the Louvain community detection algorithm. Each cell of the matrix represents the strength between two nodes in the group network. (b,e) Community structure of inhibitory nodes of the SyIPN and SPN. (c,f) Community structure of excitatory nodes of the SyIPN and SPN. Inhibitory and excitatory neural communities are visualized on three-dimensional brain renderings using BrainNet Viewer software.

with right auditory cortex and bilateral cerebellum; left cerebellum with right occipital cortex; and bilateral cerebellum with bilateral nucleus accumbens (figure 3*b*, green). The strongest SyIPN-specific connections involved the cerebellum (left lobule VII and right lobules VII, VIII), inferior occipital gyrus and nucleus accumbens.

In contrast, the SPN was characterized by distinct connections of the left prefrontal cortex with the right primary somatosensory and inferior parietal cortex, bilateral cerebellum with the bilateral prefrontal, auditory and right inferior/superior parietal cortex, hippocampus and amygdala (figure 3*b*, purple). The strongest SPN-specific connections involved the right inferior frontal and temporal cortex, left hippocampus and bilateral cerebellum (left lobule VII and bilateral lobule X).

(c) Hubs of the syllable and speech production networks

In the SyIPN, two nodes in the left inferior frontal gyrus (pars orbitalis) and medial orbitofrontal gyrus formed inhibitory hubs, while 17 nodes in the bilateral superior parietal

cortex (areas 7A, 7M, 7P, left precuneus, right areas 5L, 7PC), left superior occipital gyrus and bilateral cuneus, posterior cingulate cortex, thalamus (temporal division) and cerebellum (lobules I/IV, V, VI, Viv) formed excitatory hubs (figure 4*a,b*). Among these hubs, nine (one inhibitory; eight excitatory) were connectors, establishing connections between different neural communities, and 10 (one inhibitory; nine excitatory) were provincials, establishing connections within their own communities (figure 4*b*). The majority of SyIPN hubs were in the first neural community ($N=10$), followed by the second ($N=5$) and third ($N=4$) neural communities (figure 4*b*).

The SPN had three inhibitory hubs, including the same two inhibitory hubs of the SyIPN and an additional hub in the left cerebellar lobule X (figure 4*a,b*). The SPN excitatory hubs comprised 20 nodes, including nearly half (41%) of those present in the SyIPN—superior parietal areas 7A (left), 7P (bilateral), 5L (right), left precuneus and bilateral cuneus—in addition to left superior parietal areas 5L, 5M, 7M, 7PC, inferior parietal areas PGm and PGa, bilateral primary motor area 4a, right primary somatosensory area 2 and insula, left middle cingulate cortex, auditory areas HIP1 and HIP3, and right middle temporal

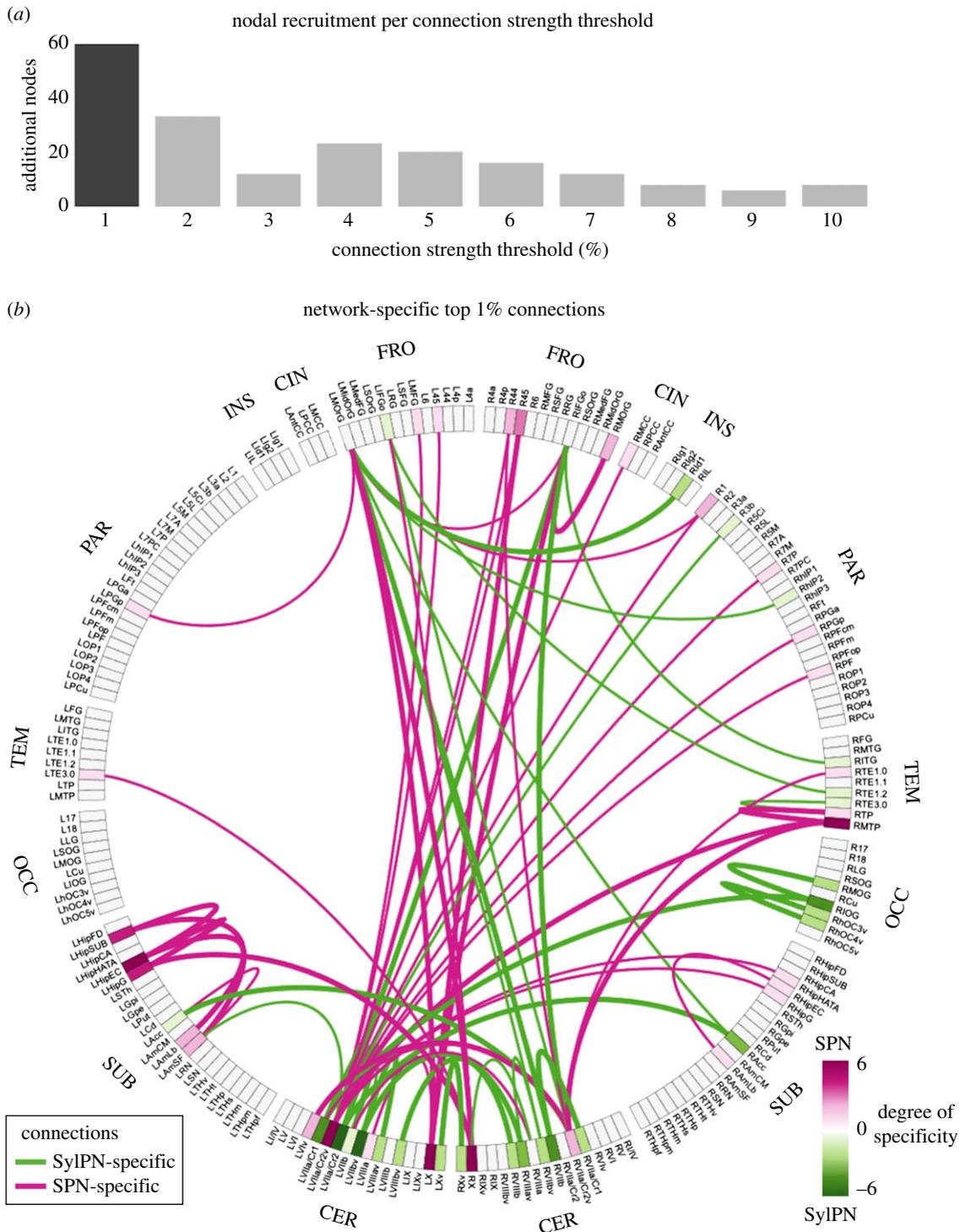


Figure 3. Network-specific most influential connections. (a) The bar graph represents the number of brain regions expressing network-specific connections recruited when increasing the strength threshold by 1% steps. The darker bar represents the most stringent connection threshold (1%) which allows recruitment of the largest number of network-specific nodes. (b) The strongest network-specific connections at 1% threshold are represented as a connectogram. Labels on the outer circle show the anatomical lobe for each node. Each brain region is indicated by an abbreviation. Inside the circle, the strongest 1% of connections exhibiting the difference between the SPN and SyIPN are displayed as edges, SPN-specific (pink) and SyIPN-specific (green). The thickness of each line represents a unidirectional (thin line) or bidirectional (thick line) connection specific to that network. The colour of each brain region label shows the degree of specificity of a given region, which represents to what extent the node is involved in SPN-specific or SyIPN-specific connections. Abbreviations: FRO = frontal; CIN = cingulate; INS = insula; PAR = parietal; TEM = temporal; OCC = occipital; SUB = subcortical; CER = cerebellum.

gyrus (figure 4*a,b*). Among these, three hubs (two inhibitory, one excitatory) were connectors linking differential neural communities, and 20 hubs (1 inhibitory, 19 excitatory) were provincials linking regions within their own neural communities (figure 4*b*). The majority of SPN hubs were found in the second neural community ($N=14$), followed by the first ($N=5$) and third ($N=4$) neural communities (figure 4*b*). Hubs in

both the SyIPN and SPN were left-hemisphere dominant, including 14 out of 19 SyIPN hubs and 16 out of 23 SPN hubs.

(i) **Effective connectivity of shared hub network of syllable production network and speech production network**

Although both the SyIPN and SPN shared nine hubs in frontal, parietal and occipital areas, their directional patterns of

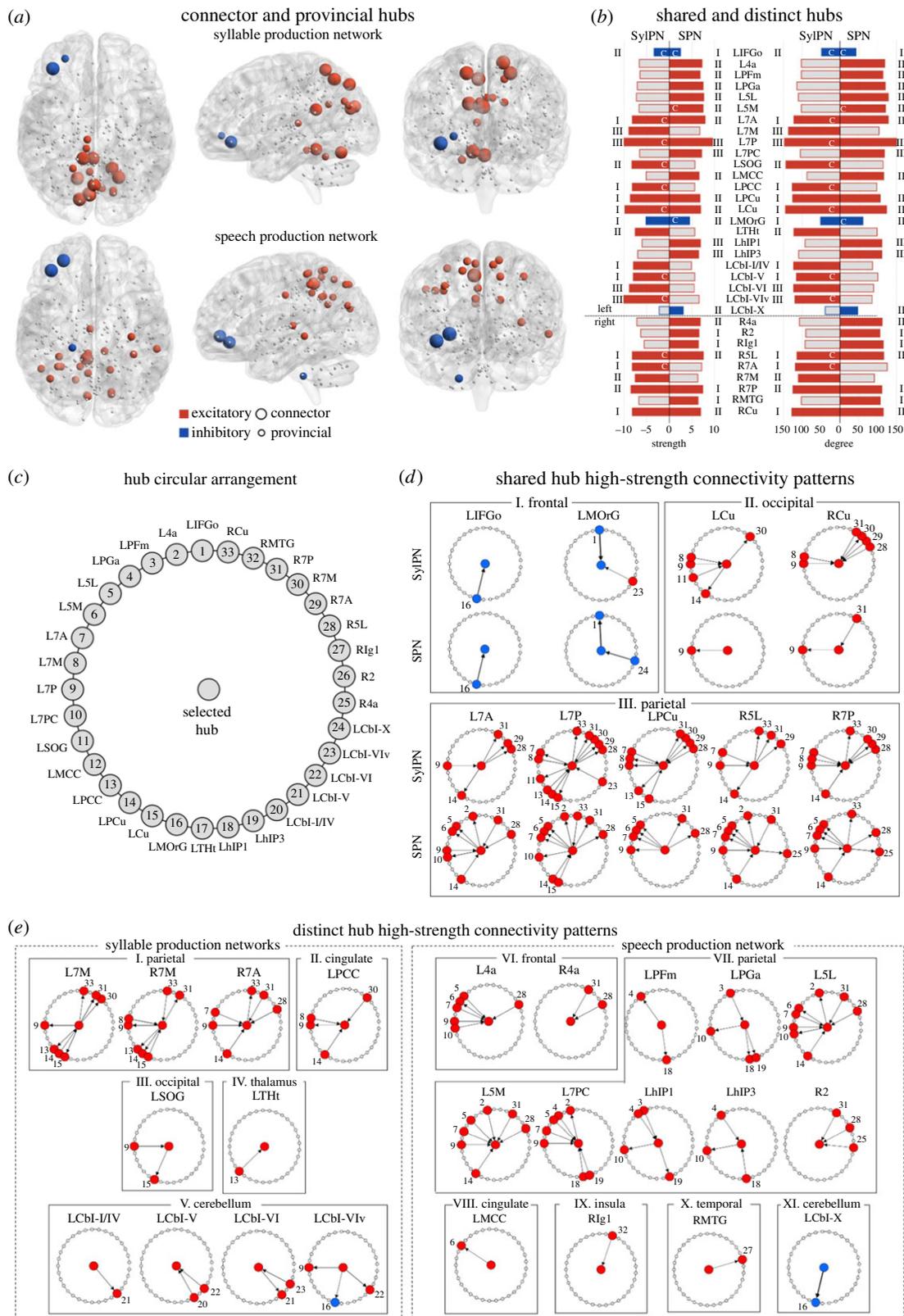


Figure 4. Hub topology of the group networks during syllable and speech production. *(a)* Inhibitory (blue) and excitatory (red) hubs in the SyIPN and SPN, defined as nodes with inner/outer strength and degree higher than one standard deviation above the average inner/outer strength and degree of each group network. Larger nodes represent connector hubs, smaller nodes represent provincial hubs. Hubs are shown on three-dimensional brain renderings using BrainNet Viewer software. *(b)* Normalized strength and degree of shared and distinct hubs in the SyIPN and SPN. The bar colour indicates excitatory (red) or inhibitory (blue) hub; grey bars indicate the absence of that region as a network hub. Bold regions indicate shared excitatory/inhibitory hubs between the networks. Numbers on the left/right side of the figure represent the modular assignment of each hub in the SyIPN/SPN—figure 2. *(c)* Schematic of hub arrangement corresponding to *(d)* and *(e)*. Each hub is assigned a corresponding number for easier identification. *(d,e)* High-strength/SPN connectivity patterns between shared *(d)* and distinct *(e)* hubs of the SyIPN and SPN. The node in the middle of each circle represents the hub, connectivity of which is displayed. Nodes on the circles represent excitatory (red) and inhibitory (blue) hubs connected to the central hub with a high-strength connection, arranged as in *(c)* and identified by the corresponding number. Abbreviations: L = left; R = right; 2, 4a, 5L, 5M, 7A, 7M, 7P, 7PC = area 2, 4a, 5L, 5M, 7A, 7M, 7P, 7PC; Cbl-I/IV, V, VI, VIv, X = cerebellar lobules I/IV, V, VI, VIv, X; Cu = cuneus; hIP1, hIP3, PFm, PGa = area hIP1, hIP3, PFm, PGa of the inferior parietal cortex; IFGo = inferior frontal gyrus; Ig1 = area Ig1 of the insula; MCC = middle cingulate cortex gyrus; MORg = medial orbital gyrus; MTG = middle temporal gyrus; PCC = posterior cingulate cortex gyrus; PCu = precuneus; SOG = superior occipital gyrus; THt = temporal area of the thalamus.

information transfer (i.e., direct influence upon another region) differed considerably when comparing syllable with speech production (figure 4c,d).

Within the frontal lobe, the left inferior frontal hub received a projection from the left cuneus in both the SylPN and SPN; however, the left medial orbitofrontal hub had an incoming projection from the left inferior frontal area in the SPN (figure 4d (I)). Moreover, incoming projections to the left medial orbitofrontal gyrus from the excitatory left cerebellar lobule VIv hub in the SylPN were switched to the incoming projection from the inhibitory left cerebellar lobule X in the SPN.

Within the occipital lobe, bilateral cuneus hubs had a broader connectivity pattern in the SylPN than in the SPN (figure 4d (II)). The left excitatory cuneus hub involved outgoing projections to right superior parietal area 7M and the left precuneus as well as incoming projections from left superior parietal areas 7M, 7P, and superior occipital gyrus. On the contrary, the left cuneus in the SPN established an outgoing connection with excitatory left superior parietal area 7P. The right cuneus hub in the SylPN had an outgoing projection to right superior parietal area 7M and incoming projections from bilateral superior parietal area 7M and right areas 5L, 7A, 7P. The directions of cuneus connectivity with superior parietal areas 7A (left) and 7P (right) were reversed in the SPN compared with the SylPN.

Finally, the most complex and dense connectivity pattern was seen for the shared parietal hubs in both the SylPN and SPN (figure 4d (III)). All parietal hubs of the SylPN established outgoing connectivity with right superior parietal area 7A, whereas all parietal hubs of the SPN had outgoing projections to the left superior parietal area 5L. Other most frequent connections were directed to bilateral superior parietal areas 7M, 7A, 7P, left precuneus and cuneus in the SylPN and left primary motor area 4a, superior parietal areas 5L, 5M, 7A, 7P and precuneus in the SPN.

(ii) Effective connectivity of distinct hub network of syllable production network and speech production network

The majority of hubs in each network were not shared with the other network (53% in SylPN, 61% in SPN). The SylPN distinctly included hubs in occipital and thalamic regions, whereas the SPN had distinct hubs in primary motor, inferior parietal, insular and middle temporal areas (figure 4a–c,e). In addition, while both the SylPN and SPN involved superior parietal, cingulate and cerebellar regions in their hub networks, their respective network hubs occupied distinct areas within these regions.

In the SylPN, parietal hubs in bilateral superior parietal area 7M and right area 7A established similar connectivity with bilateral superior parietal areas 7P and cuneus (figure 4e (I)). The left posterior cingulate hub established an outgoing projection to left superior parietal area 7P and incoming projections from bilateral superior parietal area 7M and left precuneus (figure 4e (II)). The left superior occipital hub had an outgoing projection to the left cuneus and an incoming projection from superior parietal area 7P (figure 4e (III)). The left thalamus (temporal subdivision) had an incoming projection from left posterior cingulate cortex (figure 4e (IV)), and the left cerebellar hubs in lobules I/IV, V, VI had common interlobular connections, while the left cerebellar hub in lobule VIv

had additional outgoing projections to left superior parietal area 7P (excitatory) and the left medial orbital gyrus (inhibitory) (figure 4e (V)).

In the SPN, both left and right primary motor area 4a hubs received a projection from right superior parietal area 5L, whereas left area 4a had additional incoming projections from left superior parietal areas 5L, 5M, 7A, 7P, 7PC (figure 4e (VI)). Parietal hubs were prevalent within the SPN and included left inferior parietal areas PFm, PGa, hIP1, hIP3 and superior parietal areas 5L, 5M, 7PC, with common connections within the bilateral parietal cortex and outgoing projections to left primary motor area 4a (figure 4e (VII)). In addition, the hub in right primary somatosensory area 2 received projections from the right superior parietal cortex and primary motor area 4a. The middle cingulate hub sent a projection to left superior parietal area 5M (figure 4e (VIII)), while the right insular hub and middle temporal hub established bidirectional projections (figure 4e (IX, X)) and received a projection from the right middle temporal gyrus (figure 4e (IX)). The cerebellum contributed the only inhibitory hub in the left cerebellar lobule X, which sent an outgoing projection to the left middle orbital gyrus within the hub network (figure 4c (VI)).

(iii) Long-range influence of network hubs

The projection system and extent of each hub involved in both the SylPN and SPN were assessed by computing the distance of the farthest node with an excitatory or inhibitory connection to the hub. The SPN hubs projected to and influenced brain regions at a greater distance than SylPN hubs (projection distance: 105.4 ± 16.3 mm in SPN; 101.4 ± 17.0 mm in SylPN; $p = 0.049$, Wilcoxon signed-rank test) by establishing longest ranging projections with left prefrontal and bilateral superior/inferior regions (figure 5).

4. Discussion

Using a computationally efficient rDCM and graph-theoretical analyses of fMRI data in healthy individuals, we examined the whole-brain dynamic connectome during the production of meaningful, real-life speech compared with a meaningless vocal motor task, syllable. In line with our previous study, which described the functional connectome of speech control [12], the current findings demonstrate that speech production involves a highly complex orchestration of brain regional connectivity. Here, our model-based approach further inferred information about directional patterns of information transfer within speech and syllable production networks, allowing, for the first time to our knowledge, expansion of our understanding of directed regional influences within these specialized large-scale connectomes. Specifically, we determined that speaking is regulated by a directed neural network, the most influential nodes of which are centred around primary sensorimotor and parietal cortical areas and preferentially influence prefrontal regions via long-ranging functional connectivity. Below, we discuss the dynamic connectivity of the SPN in detail, with comparisons to the SylPN.

The SPN architecture was characterized by a more segregated network topology than the SylPN. The SPN included four neural communities, as opposed to three communities of the SylPN (figure 2), which was achieved by forming smaller but stronger integrated populations of brain regions, pointing to the greater specialization of the SPN function. Notably, the additional fourth SPN community was

short- and long-distance influence of network hubs

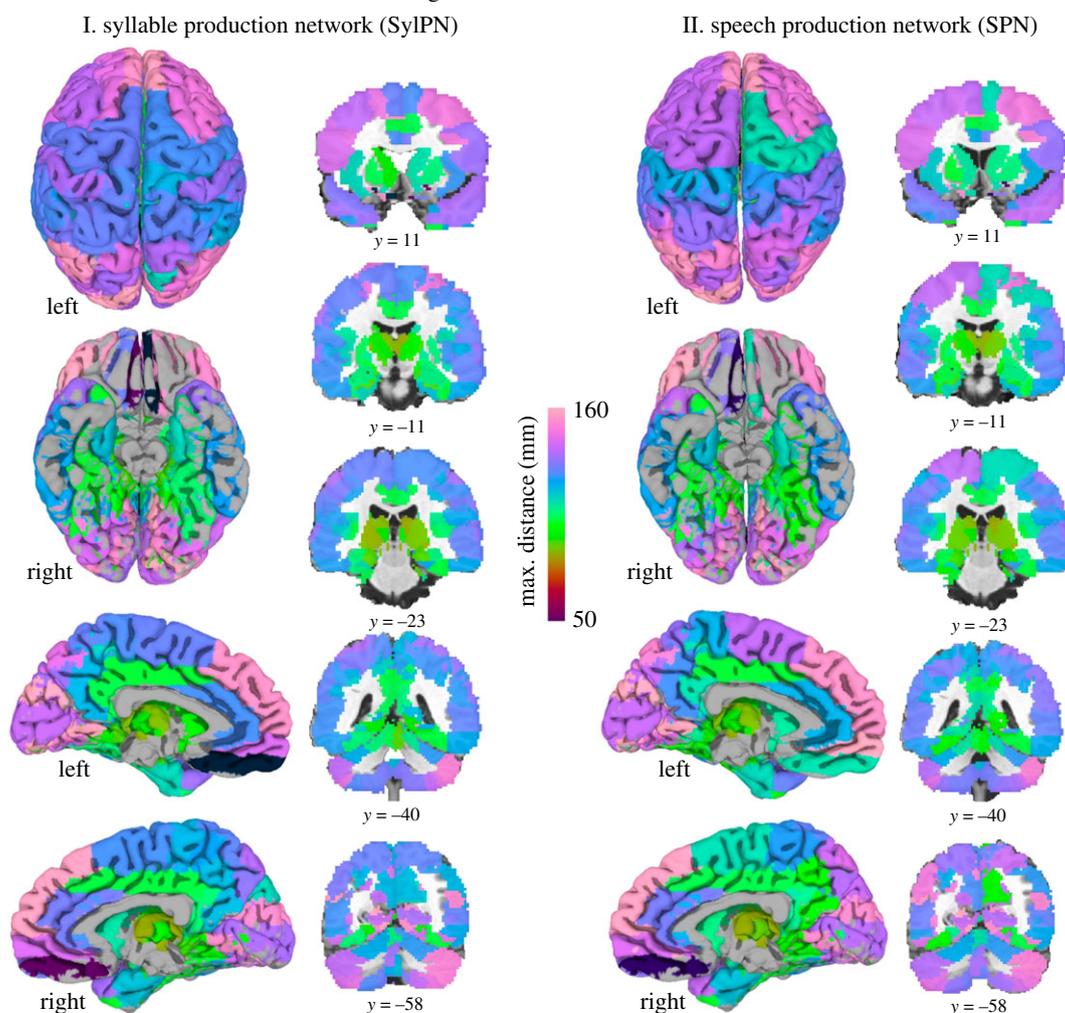


Figure 5. Distance influence of network hubs in the SyIPN and SPN. Spatial influence of hubs in the SyIPN (I) and SPN (II), defined by the farthest node to which they are connected. Colours represent the Euclidean distance between each hub and the farthest node to which they are connected, measured in millimetres. ROIs are visualized using AFNI/SUMA software.

composed of brain regions involved in various aspects of speech control, including visuospatial transformations for integration with motor action planning (intra-parietal area hIP2) [29], semantic verbal fluency (lateral globus pallidus) [30], generation of pre-articulatory verbal code (cerebellar lobule VI) [31], auditory prosodic categorization (amygdala) [32], and articulo-phonological processing (red nucleus) [33]. On the other hand, the SPN established only three connector hubs, which interconnected its four neural communities, versus 20 provincial hubs that exerted influence within their own subcommunities (figure 4). These data suggest that fewer connector hubs, including left inferior frontal, medial orbitofrontal and superior parietal cortex, provide the multi-modal integration of the SPN, whereas a prevailing number of provincial hubs distributed across different subdivisions of primary sensorimotor, parietal, cingulate, insular, temporal, occipital and cerebellar regions support network activity through controlling smaller groups of neural populations. Conversely, the SyIPN hubs were nearly equally split between connectors and provincials (9 versus 10), indicating a more uniform distribution of the overall network regulatory influence between and within communities.

Given that syllables are building blocks of connected speech production, the SyIPN and SPN shared the same set of hubs in prefrontal, parietal and occipital areas (figure 4b).

However, their respective directional connectivity profiles differed substantially between the two networks (figure 4d). For example, the SPN hub in left superior parietal area 7A, which is known to be involved in motor sequencing of speech production [31], had wider-reaching incoming and outgoing connections with other bilateral parietal areas and left primary motor area 4a. The same hub in the SyIPN had predominantly outgoing connections that were limited within the superior parietal cortex. By contrast, hubs in the bilateral cuneus, which is known to be active during listening, production and reading of the word lists [32], had a more enhanced projection profile, connecting to six different subdivisions of the surrounding parietal cortex in the SyIPN. The same hubs in the SPN were restricted to only one outgoing and one incoming parietal projection. These findings suggest that each network is able to reorganize its hubs for prioritization of information transfer that is specifically associated with the complexity of a given task production.

The majority of shared hubs in both networks were excitatory, influencing or being influenced by other excitatory regions. Only two hubs were inhibitory, including the left inferior frontal gyrus pars orbitalis, known for its role in lexical retrieval and emotional prosody [33,34], and the left medial orbitofrontal gyrus, involved in error phoneme categorization [35], reflecting their essential modulatory function in supporting both basic and complex vocal motor production.

The most notable differences between the SyIPN and SPN were observed in the organization of the behaviour-specific hubs. Specifically, the SyIPN had a distinct overrepresentation of excitatory cerebellar hubs (figure 4e). The cerebellum is involved in the modulation of syllabic timing and rhythmic structure, concatenation of syllable strings into articulated sentences, and error-driven adjustment of motor commands [36–38]. Cerebellar hubs participated in bidirectional information transfer between different cerebellar lobules and directly influenced orbitofrontal and superior parietal activity, thus acting at both regional and large-scale SyIPN levels. It is notable that the thalamus was another distinct hub of the SyIPN, as it lies on the path of cerebellar connectivity with cortical regions. The presence of the cerebellum and thalamus as distinct SyIPN hubs likely allows enhanced timing precision during repetitive syllable production. The absence of the thalamus and several cerebellar regions from the SPN hubs may be due to the increased complexity of the speech task, which prioritized the involvement of other brain regions as hubs over the thalamus and cerebellum.

In contrast to the SyIPN, the SPN recruited primary motor and somatosensory hubs that were influenced by heavy projections from bilateral superior parietal areas, inferior parietal hubs that further broadened the overall influence of intra-parietal connectivity, and bidirectionally connected middle temporal and insular hubs (figure 4e). The presence of these distinct SPN hubs and their respective directional connectivity profiles likely allows the balanced integration of multimodal processes necessary for speech perception and processing, sensorimotor transformations, and motor output. While the roles of the inferior frontal gyrus and primary motor cortex are at the centre of discussion when considering neural control of speech production [12,39–43], our data show that the major information flow within the SPN is in fact directed through the superior/inferior parietal cortex, which subsequently influences other speech-associated cortical areas and seemingly integrates them into the neural network for speech production. Notably, in the SPN compared with the SyIPN, prefrontal areas, including the inferior frontal gyrus, appear to receive denser (figure 3) and longer-ranging

(figure 5) projections from parietal and cerebellar regions that likely influence and contribute to more refined speech monitoring and motor preparation.

In summary, we combined fMRI data with a novel computational approach of rDCM and graph-theoretical analysis to examine whole-brain effective connectivity during real-life speaking versus simple syllable production. We demonstrated that neural network undergoes reorganization of its connectivity with increased complexity during speaking. Highly influential hubs in superior and inferior parietal areas influence the whole-brain network activity via directed and long-ranging projections with primary sensorimotor, prefrontal, temporal and insular areas, which together support the multimodal information processing for speech motor output. By contrast, the neural network during simpler syllable production is characterized by a more compressed, cost-efficient structure, which supports the essential elements of sequence timing and sensorimotor integration via the influence exerted by superior parietal and cerebellar network hubs.

Ethics. All participants signed informed consent before taking part in the study, which was approved by the Institutional Review Board of Massachusetts General Brigham (protocol no. 2019P001576) and Icahn School of Medicine at Mount Sinai (protocol no. 10-1362).

Data accessibility. Data and codes are available at <https://github.com/simonyanlab/Analytic-Tools>.

Authors' contributions. D.V. participated in the design of the study, performed data and statistical analyses, and drafted the manuscript. K.S. participated in the design of the study, collected data, coordinated the study, and critically revised the manuscript. All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. D.V. declares no competing financial interests. K.S. serves on the Scientific Advisory Board of the Tourette Association of America.

Funding. This research was supported by the National Institute on Deafness and Other Communication Disorders and National Institutes of Health (grant no. R01DC011805).

Acknowledgements. The authors would like to thank Stefan Frässle, PhD, for his support in the adaptation of rDCM pipeline for sparse-sampled event-related fMRI data, and Azadeh Hamzehei-Sichani, MA, for assistance with subject recruitment and MRI data acquisition.

References

1. Simonyan K, Ostuni J, Ludlow CL, Horwitz B. 2009 Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *J. Neurosci.* **29**, 14 912–14 923. (doi:10.1523/JNEUROSCI.4897-09.2009)
2. Korzeniewska A, Franaszczuk PJ, Crainiceanu CM, Kus R, Crone NE. 2011 Dynamics of large-scale cortical interactions at high gamma frequencies during word production: event related causality (ERC) analysis of human electrocorticography (ECoG). *Neuroimage* **56**, 2218–2237. (doi:10.1016/j.neuroimage.2011.03.030)
3. Riecker A, Mathiak K, Wildgruber D, Erb M, Hertrich I, Grodd W, Ackermann H. 2005 fMRI reveals two distinct cerebral networks subserving speech motor control. *Neurology* **64**, 700–706. (doi:10.1212/01.WNL.0000152156.90779.89)
4. Peeva MG, Guenther FH, Tourville JA, Nieto-Castanon A, Anton JL, Nazarian B, Alario FX. 2010 Distinct representations of phonemes, syllables, and supra-syllabic sequences in the speech production network. *Neuroimage* **50**, 626–638. (doi:10.1016/j.neuroimage.2009.12.065)
5. Fu CH, McIntosh AR, Kim J, Chau W, Bullmore ET, Williams SC, Honey GD, McGuire PK. 2006 Modulation of effective connectivity by cognitive demand in phonological verbal fluency. *Neuroimage* **30**, 266–271. (doi:10.1016/j.neuroimage.2005.09.035)
6. Weber S, Hausmann M, Kane P, Weis S. 2020 The relationship between language ability and brain activity across language processes and modalities. *Neuropsychologia* **146**, 107536. (doi:10.1016/j.neuropsychologia.2020.107536)
7. Heim S, Opitz B, Müller K, Friederici AD. 2003 Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Brain Res. Cogn. Brain Res.* **16**, 285–296. (doi:10.1016/S0926-6410(02)00284-7)
8. Prabhakaran R, Blumstein SE, Myers EB, Hutchison E, Britton B. 2006 An event-related fMRI investigation of phonological–lexical competition. *Neuropsychologia* **44**, 2209–2221. (doi:10.1016/j.neuropsychologia.2006.05.025)
9. Hickok G, Houde J, Rong F. 2011 Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* **69**, 407–422. (doi:10.1016/j.neuron.2011.01.019)
10. Simmonds AJ, Wise RJ, Collins C, Redjep O, Sharp DJ, Iverson P, Leech R. 2014 Parallel systems in the control of speech. *Hum. Brain Mapp.* **35**, 1930–1943. (doi:10.1002/hbm.22303)
11. de Pasquale F, Della Penna S, Snyder AZ, Marzetti L, Pizzella V, Romani GL, Corbetta M. 2012 A cortical

- core for dynamic integration of functional networks in the resting human brain. *Neuron* **74**, 753–764. (doi:10.1016/j.neuron.2012.03.031)
12. Fuertinger S, Horwitz B, Simonyan K. 2015 The functional connectome of speech control. *PLoS Biol.* **13**, e1002209. (doi:10.1371/journal.pbio.1002209)
 13. Simonyan K, Fuertinger S. 2015 Speech networks at rest and in action: interactions between functional brain networks controlling speech production. *J. Neurophysiol.* **113**, 2967–2978. (doi:10.1152/jn.00964.2014)
 14. Fuertinger S, Simonyan K. 2016 Stability of network communities as a function of task complexity. *J. Cogn. Neurosci.* **28**, 2030–2043. (doi:10.1162/jocn_a_01026)
 15. Friston KJ, Harrison L, Penny W. 2003 Dynamic causal modelling. *Neuroimage* **19**, 1273–1302. (doi:10.1016/s1053-8119(03)00202-7)
 16. Frassle S, Lomakina EI, Razi A, Friston KJ, Buhmann JM, Stephan KE. 2017 Regression DCM for fMRI. *Neuroimage* **155**, 406–421. (doi:10.1016/j.neuroimage.2017.02.090)
 17. de Lima Xavier L, Simonyan K. 2020 Neural representations of the voice tremor spectrum. *Mov. Disord.* **35**, 2290–2300. (doi:10.1002/mds.28259)
 18. Eickhoff SB, Stephan KE, Mohlberg H, Grefkes C, Fink GR, Amunts K, Zilles K. 2005 A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *Neuroimage* **25**, 1325–1335. (doi:10.1016/j.neuroimage.2004.12.034)
 19. Friston K, Moran R, Seth AK. 2013 Analysing connectivity with Granger causality and dynamic causal modelling. *Curr. Opin. Neurobiol.* **23**, 172–178. (doi:10.1016/j.conb.2012.11.010)
 20. Frässle S *et al.* 2021 Regression dynamic causal modeling for resting-state fMRI. *Hum. Brain Mapp.* **42**, 2159–2180. (doi:10.1002/hbm.25357)
 21. Lynall ME, Bassett DS, Kerwin R, McKenna PJ, Kitzbichler M, Muller U, Bullmore E. 2010 Functional connectivity and brain networks in schizophrenia. *J. Neurosci.* **30**, 9477–9487. (doi:10.1523/JNEUROSCI.0333-10.2010)
 22. Frassle S, Lomakina EI, Kasper L, Manjaly ZM, Leff A, Pruessmann KP, Buhmann JM, Stephan KE. 2018 A generative model of whole-brain effective connectivity. *Neuroimage* **179**, 505–529. (doi:10.1016/j.neuroimage.2018.05.058)
 23. Fagiolo G. 2007 Clustering in complex directed networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **76**, 026107. (doi:10.1103/PhysRevE.76.026107)
 24. Betzel RF, Griffa A, Hagmann P, Misisic B. 2019 Distance-dependent consensus thresholds for generating group-representative structural brain networks. *Network Neurosci.* **3**, 475–496. (doi:10.1162/netn_a_00075)
 25. Newman ME. 2006 Modularity and community structure in networks. *Proc. Natl Acad. Sci. USA* **103**, 8577–8582. (doi:10.1073/pnas.0601602103)
 26. Sun Y, Danila B, Josić K, Bassler KE. 2009 Improved community structure detection using a modified fine-tuning strategy. *EPL* **86**, 28004. (doi:10.1209/0295-5075/86/28004)
 27. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008 Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008. (doi:10.1088/1742-5468/2008/10/P10008)
 28. Rubinov M, Sporns O. 2010 Complex network measures of brain connectivity: uses and interpretations. *Neuroimage* **52**, 1059–1069. (doi:10.1016/j.neuroimage.2009.10.003)
 29. Hagberg A, Swart P, Chult SD. 2008 *Exploring network structure, dynamics, and function using NetworkX*. Los Alamos, NM: Los Alamos National Laboratory.
 30. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009 Circos: an information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645. (doi:10.1101/gr.092759.109)
 31. Heim S, Amunts K, Hensel T, Grande M, Huber W, Binkofski F, Eickhoff SB. 2012 The role of human parietal area 7A as a link between sequencing in hand actions and in overt speech production. *Front. Psychol.* **3**, 534. (doi:10.3389/fpsyg.2012.00534)
 32. Hesling I, Labache L, Joliot M, Tzourio-Mazoyer N. 2019 Large-scale plurimodal networks common to listening to, producing and reading word lists: an fMRI study combining task-induced activation and intrinsic connectivity in 144 right-handers. *Brain Struct. Funct.* **224**, 3075–3094. (doi:10.1007/s00429-019-01951-4)
 33. Seydell-Greenwald A, Chambers CE, Ferrara K, Newport EL. 2020 What you say versus how you say it: comparing sentence comprehension and emotional prosody processing using fMRI. *Neuroimage* **209**, 116509. (doi:10.1016/j.neuroimage.2019.116509)
 34. Conner CR, Kadipasaoglu CM, Shouval HZ, Hickok G, Tandon N. 2019 Network dynamics of Broca's area during word selection. *PLoS ONE* **14**, e0225756. (doi:10.1371/journal.pone.0225756)
 35. Blanco-Elorrieta E, Gwilliams L, Marantz A, Pykkänen L. 2021 Adaptation to mis-pronounced speech: evidence for a prefrontal-cortex repair mechanism. *Scient. Rep.* **11**, 97. (doi:10.1038/s41598-020-79640-0)
 36. Ackermann H, Mathiak K, Riecker A. 2007 The contribution of the cerebellum to speech production and speech perception: clinical and functional imaging data. *Cerebellum* **6**, 202–213. (doi:10.1080/14734220701266742)
 37. Simonyan K, Ackermann H, Chang EF, Greenlee JD. 2016 New developments in understanding the complexity of human speech production. *J. Neurosci.* **36**, 11 440–11 448. (doi:10.1523/JNEUROSCI.2424-16.2016)
 38. Manto M *et al.* 2012 Consensus paper: roles of the cerebellum in motor control—the diversity of ideas on cerebellar involvement in movement. *Cerebellum* **11**, 457–487. (doi:10.1007/s12311-011-0331-9)
 39. Price CJ. 2012 A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* **62**, 816–847. (doi:10.1016/j.neuroimage.2012.04.062)
 40. Ardila A, Bernal B, Rosselli M. 2016 How localized are language brain areas? A review of Brodmann areas involvement in oral language. *Arch. Clin. Neuropsychol.* **31**, 112–122. (doi:10.1093/arclin/acv081)
 41. Scott SK. 2012 The neurobiology of speech perception and production—can functional imaging tell us anything we did not already know? *J. Commun. Disord.* **45**, 419–425. (doi:10.1016/j.jcomdis.2012.06.007)
 42. Hickok G. 2012 Computational neuroanatomy of speech production. *Nat. Rev. Neurosci.* **13**, 135–145. (doi:10.1038/nrn3158)
 43. Guenther FH. 2016 *Neural control of speech*. Cambridge, MA: MIT Press.