bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Neural representations of speech production in neocortical and cerebellar regions

Sivan Jossinger[1,2*], Bassel Arafat[1,2], Jorn Diedrichsen[1-3]

[1]Department of Computer Science, Western University, London Ontario, Canada, N6A 3K7

[2]Western Center of Brain and Mind, Western University, London Ontario, Canada, N6A 3K7

[3]Department of Statistical and Acturial Sciences, Western University, London Ontario, Canada, N6A 3K7

*Corresponding author: Sivan Jossinger (sjossing@uwo.ca)

Abstract

Speech production depends on the precise temporal integration of articulatory movements with phonation. While ventral primary motor cortex is known to encode articulatory features, how phonatory timing, and its coordination with articulation, is represented across cortical and cerebellar circuits remains poorly understood. Using 7T functional MRI, we examined neural representations during overt syllable production varying in place of articulation and voice onset time. Multivariate analyses revealed reliable, syllable-specific differences in activity patterns across both cortical and cerebellar speech regions. Ventral primary sensorimotor cortex distinguished syllables by place of articulation, whereas dorsal sensorimotor cortex was more sensitive to the timing of voice onset relative to articulation. Secondary sensorimotor speech areas, including the operculum and auditory cortex, showed a hybrid representational profile, integrating both articulatory and phonatory features. In the cerebellum, representational geometry was dominated by the place of articulation; however, overall syllable representations were most similar to those in the operculum, accounting for unique variance beyond that explained by ventral sensorimotor cortex. Together, these findings reveal feature-specific representational tuning across primary sensorimotor regions during speech production. The selective representational alignment between operculum and cerebellum may support the refinement of speech motor plans prior to execution.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Speech is among the most complex motor activities humans perform, requiring the coordination of approximately 100 muscles across laryngeal, respiratory and oral motor systems. This coordination integrates two fundamental speech features: articulation, which shapes the configuration of the vocal tract, and phonation, the generation of voiced sound. Such intricate orchestration depends on a distributed neural network spanning both cortical and cerebellar regions (Guenther & Hickok, 2016). Here, we aim to determine how articulatory and phonatory features of speech are represented across cortical and cerebellar regions during syllable production.

Articulatory control is primarily attributed to the ventral sensorimotor cortex (vSM), where upper motor neurons project to brainstem nuclei, which in turn innervate the vocal tract via various cranial nerves (Jürgens, 2002; Penfield & Boldrey, 1937). Electrophysiological and imaging studies have shown that neural populations in vSM are selectively tuned to specific articulators (lips, tongue, and jaw), following a somatotopic layout that recapitulates the vocal tract (Bouchard et al., 2013; Carey et al., 2017). However, how speech features beyond articulation, such as phonation, are represented remains unclear.

Phonation is the process of generating sound through the vocal folds (i.e., voicing). This process involves steady exhalation and laryngeal muscle contraction to vibrate the vocal folds. Cortically, laryngeal control has been mapped to two distinct regions in the vSM (Bouchard et al., 2013; Eichert et al., 2020), while exhalation has been shown to recruit a specific region in the dorsal sensorimotor cortex (dSM) associated with trunk movement (Brown et al., 2009; Loucks et al., 2007). Notably, the dSM shows greater activation during voiced compared with whispered speech, suggesting a role in phonatory control that is not limited to respiration (Correia et al., 2020).

Beyond the primary sensorimotor cortex, clinical evidence highlights a critical role for the cerebellum in speech production (Ackermann & Brendel, 2016). Cerebellar motor regions are embedded in closed-loop circuits with the cerebral sensorimotor cortex. In these circuits, individual cerebellar motor territories are reciprocally connected with specific, somatotopically organized cortical motor areas (Kelly & Strick, 2003; Saadon-Grosman et al., 2022). This closed-loop cerebellar-cortical organization predicts that cerebellar representations should closely resemble those in vSM. However, recent evidence suggests that the cerebellum integrates signals

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

from multiple cortical sources (King et al., 2023). Under this integrative account, cerebellar speech regions may exhibit mixed representational geometries. Specifically, these regions may combine articulatory and phonatory features within their neural activity patterns.

In the current study, we used high-field (7T) fMRI to investigate the neural mechanisms underlying speech production. Participants uttered syllables that differed in place of articulation and voicing. Leveraging the high spatial resolution afforded by 7T imaging, we aimed to map cortical and cerebellar regions associated with speech production. We employed Representational Similarity Analysis (RSA) to determine how specific speech features are encoded in neural activity patterns. Activity patterns in vSM primarily distinguished between places of articulation, consistent with previous work (Bouchard et al., 2013). Representational geometries in dSM differentiated between voiced and voiceless consonants, in line with the involvement of this area in phonation (Correia et al., 2020). Secondary sensorimotor speech areas outside primary motor cortex showed a mixture of these two representations. We then tested whether speech representations in the cerebellum resemble those observed in cortical speech regions. We hypothesized that cerebellar speech regions would show representational structures similar to those in vSM, consistent with closed-loop cerebellar-cortical organization (Kelly & Strick, 2003; Saadon-Grosman et al., 2022). Alternatively, cerebellar speech regions may encode mixed representational geometries that combine ventral and dorsal sensorimotor features, reflecting the convergence of cortical inputs to the cerebellum (King et al., 2023).

Methods

Participants. Twelve neurotypical adults were recruited for this study (6 females, 18-29 years [mean ± SD = 23.3 ± 3.6]). All participants were right handed as estimated by the Edinburgh handedness inventory (90.7 ± 15.1; Oldfield, 1971), native-level English speakers, and had no history of speech impairment or a neurological condition. All experimental procedures were approved by the Research Ethics Committee at Western University. The participants signed a written informed consent before participating in the study and were compensated for their participation.

Stimuli. Six different consonant-vowel (CV) syllables were visually presented to the participants at the center of the screen using PsychoPy (https://www.psychopy.org/). The

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

syllables were composed of a plosive consonant (/p/, /b/, /t/, /d/, /k/, /g/) followed by the vowel /a/ (Fig. 1A). Plosives are produced by blocking the airflow in the vocal tract and then releasing it, creating a burst of air. The plosive syllables in the current experiment varied along two axes: place of articulation (PoA) and voice onset time (VOT). PoA refers to the location in the vocal tract where the airflow is obstructed. In plosive sounds, the blockage can be formed with the lips (/p/, /b/), with the tongue tip against the alveolar ridge (/t/, /d/), or with the back of the tongue against the velum (/k/, /g/), corresponding to the *bilabial*, *alveolar*, and *velar* sounds, respectively. VOT refers to the interval between the release of the plosive closure and the onset of vocal fold vibration. In *voiceless* plosives (/p/, /t/, /k/), the vocal folds do not vibrate during the release, whereas in *voiced* plosives (/b/, /d/, /g/), the vocal folds do vibrate during the release.

Experimental design. Participants underwent MRI scanning in a single session. We acquired an anatomical image and 10 functional runs. Right before the MRI scan, participants were familiarized with the behavioral task in a short behavioral session (2 runs, ~20 min). On each trial, a syllable was presented on the screen for 2 seconds, followed by a fixation cross for 2 seconds (Fig. 1B). Participants were instructed to repeat the syllable out load three times at a comfortable pace during the 2s presentation period, and to remain silent during fixation to prevent overlap across task phases. Each run lasted ~4 minutes and included four blocks of 48 s separated by 14 s rest. Within each block all six syllables were presented, with each syllable appearing twice on consecutive trials. The order of items pairs within a block was randomized, resulting in a total of eight presentations of each item per run. Item and trial repetitions were included to improve the contrast-to-noise ratio (CNR). A period of 10 s rest was added at the end and of each functional run to allow for signal relaxation and provide a better estimate of baseline activation. The entire MRI session, including the anatomical scans and setup, lasted ~60 minutes.

Imaging data acquisition. Functional MRI data were acquired on a 7T Siemens Magnetom scanner with a 32-channel head coil at Western University. Anatomical T1 weighted scan of each participant was acquired at the beginning of the MRI session, using a magnetization-prepared rapid gradient echo sequence (MP2RAGE, voxel size=0.7mm isotropic; TR=6000 ms; TE=2.27 ms; field of view=246×246; 224 slices). Task-based functional data were acquired using a multi-band gradient-echo EPI sequence with anterior to posterior phase-encoding direction (voxel size= 2.3 mm isotropic; TR=1100 ms; TE=20 ms; flip angle=30; multiband acceleration

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

factor=2; GRAPPA acceleration=3; field of view=208×208 mm; 56 slices). Each run consisted of 224 volumes. To correct spatial distortions caused by inhomogeneities in the magnetic field, we also acquired a gradient-echo field map (voxel size=1.3 × 1.3 × 2.5 mm; field of view=210×210).

Preprocessing. Functional data were preprocessed in native space for each individual separately using SPM12 (fil.ion.ucl.ac.uk/spm) and custom Matlab code. Our minimal preprocessing pipeline included the following steps: First, functional images were corrected for geometric distortions caused by magnetic field inhomogeneity using the gradient echo field map (Hutton et al., 2002). Then, functional images were realigned to the first volume of the first run to correct for head motion (six parameters: translation x, y, and z, and rotation pitch, roll and yaw). Lastly, the biased-corrected functional data were co-registered to the anatomical T1 image, for which the

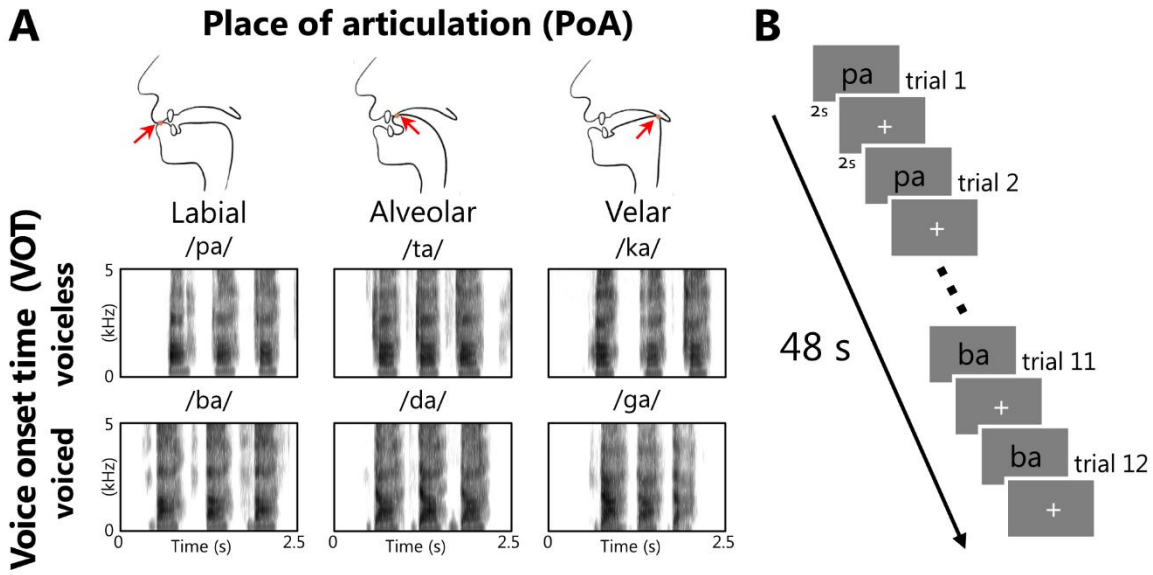

Figure 1. **Syllable repetition task**. ***A***, Experimental stimuli. Top: Schematic illustration of the vocal tract during articulation of bilabial (/p/,/b/), alveolar (/t/,/d/), and velar (/k/,/g/) plosive consonants. Red arrow denoting the place of articulation. Bottom: Spectrograms of spoken CV syllables of one representative subject (female, 18yo) recorded during the behavioral training session. Spectrograms are grouped by voice onset time, with voiceless plosives (/pa/,/ta/,/ka/) shown in the top row and voiced plosives (/ba/,/da/,/ga/) in the bottom row. ***B***, Block example. Each trial consisted of a CV syllable presented for 2 s, followed by 2 s fixation cross. Participants (N=12) were instructed to repeat the syllable three times during its presentation. Within each block, all six syllables were presented in a random order, with each syllable repeated twice on consecutive trials. Each block lasted ~48 s. Abbreviations: CV – consonant vowel; PoA – place of articulation; VOT – voice onset time.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(0,0,0) coordinate was moved to the anterior commissure (AC). No smoothing or normalization to a group template was performed at this stage.

First level general linear model. The preprocessed functional images were analyzed with a general linear model (GLM), using a separate regressor for each syllable (/p/, /b/, /t/, /d/, /k/, /g/), for each run. The activation of each trial (consisting of three repetitions of the same syllable) was modeled using a boxcar function of length 2 sec convolved with a two-gamma canonical hemodynamic response function with a peak at 5 sec and a post-stimulus undershoot minimum at 11 seconds. This analysis resulted in activation images (beta maps) for each condition per run, for each participant. Rest was not modeled explicitly but served as an implicit baseline.

Neocortical surface reconstruction. Reconstruction of cortical surface from the anatomical image was carried out using Freesurfer (Fischl et al., 1999). In this procedure, white-gray matter and pial surfaces were reconstructed for each individual. The surfaces were then inflated into a sphere, and aligned to the left-right symmetric template atlas (fs_LR.32k.spec; Van Essen et al., 2012) based on sulcal depth and curvature information. The functional data were projected from native space to the subject's individual surface, by averaging beta values of voxels intersecting the line connecting corresponding vertices of the individual white matter and pial surfaces.

Cerebellar normalization. Cerebellar isolation and segmentation into white and gray matter were performed using the Spatially Unbiased Infratentorial Template (SUIT) toolbox implemented in SPM12 (Diedrichsen, 2006). For each subject, the automatic segmentation was carefully inspected and, when necessary, manually corrected by one of the authors (S.J.) to exclude voxels originating from non-cerebellar tissue (e.g., visual cortex). Cerebellar gray and white matter maps were then normalized into SUIT space using a non-linear deformation algorithm (Ashburner, 2007). The activation estimates (i.e., beta weights) and residual mean-square from the first level GLM were also resliced into SUIT. The functional data were further resliced into MNIsymC template atlas (a symmetric version of the cerebellar only template, aligned to the MNINlin2009cSym template) to enable the definition of symmetric cerebellar ROIs in the left and right hemisphere. For visualization purposes, functional maps were projected onto a flat representation of the cerebellum using the SUIT toolbox (Diedrichsen & Zotow, 2015).

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Defining functional regions of interest. To identify brain regions activated during overt syllable repetition we generated group-level activation maps across all syllable conditions (Fig. 2). Individual subject data were projected into group space, spatially smoothed on the cortical surface or in the volume (kernel width: 6mm), and averaged across participants. To obtain a hemispheric unbiased estimate we further averaged the data across hemispheres. Functional ROIs were defined by thresholding the resulting group-average map to retain the top 10% of vertices in the neocortex, and top 10% of voxels in the cerebellum. For cerebellar ROIs, we extracted SPM t-values, reflecting average beta maps divided by the standard deviation of the residual time series at each voxel. For the neo-cortex, anatomical locations of the regions were identified using the Glasser et al. (2016) atlas.

Quantifying pattern reliability. Pattern reliability was defined as the proportion of total variance in fMRI activity patterns that could be explained by reliable effects, computed as the sum of group-level and subject-specific variance components divided by the total variance. To estimate these components, we applied variance decomposition (https://functional-fusion.readthedocs.io/) to unsmoothed activity patterns in group space (cortical or cerebellar). For each voxel or vertex and each run, we first subtracted the mean response across all syllables. This step removes shared global activations that are not specific to individual syllables. The resulting fMRI activity was decomposed into three variance components: (1) *group*: reflecting patterns that are shared across subjects; (2) *subject*: capturing reliable idiosyncratic differences; and (3) *noise*: representing run-by-run variability of the estimates within each person. These components were estimated from the covariance matrix of activity estimates across voxels, where $x_{s,i}$ denotes the pattern for subject s, in run i. The average covariance across different people, $cov(x_{s,i}, x_{t,j})$, is equal to the group variance; the average covariance within a person across runs, $cov(x_{s,i}, x_{s,j})$, is equal to the sum of group and subject variance; and the average total average variance, $var(x_{s,i})$, is equal to all three components. Pattern reliability is then the sum of the group and subject variance components divided by the total variance. Statistical significance of reliability estimates was assessed using one-sample t-tests against zero.

To compare reliability between the cortex and the cerebellum at the subject level, we conducted a linear mixed-effects model with normalized reliability as the dependent variable, brain region (i.e., cortex vs. cerebellum) as a fixed effect, and subject as a random intercept. The

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

model was fitted using Restricted Maximum Likelihood (REML). Significance was assessed with Wald z-tests.

Group-level univariate analysis. To assess whether mean activation differed systematically as a function of articulatory features, we performed a repeated-measures ANOVA on percent signal change values extracted from each ROI. The model included within-subject factors of Place of Articulation (bilabial, alveolar, velar), and Voicing (voiced, voiceless), with subject treated as a random effect. Significant main effects were followed up with post hoc comparisons, corrected for multiple comparisons (FDR correction).

To assess the topological organization of different places of articulation, we computed subject-level contrasts for each articulatory category (bilabial, alveolar, velar) against the remaining syllables. Group-level effects were then estimated by performing a one-sample t-test across participants at each vertex/voxel, yielding group t-statistic maps. These maps were used to evaluate the spatial distribution of place-of-articulation selectivity.

Multivariate pattern analysis of syllable-specific representations. To quantify how much activation patterns for each syllable differed from each other, we used the cross-validated Mahalanobis distance (Nili et al., 2014), resulting in a representational dissimilarity matrix (RDM). Prior to calculating the distances, beta weights were spatially prewhitened (i.e., scaling each voxel's beta by the estimated noise standard deviation from the GLM). This step ensures that the distance estimates reflect true pattern differences rather than noise covariance (Bosch et al., 2025). To get a cross-validated estimation of the distances, we multiplied the difference between the activity pattern of two syllables in one imaging run with the differences computed on any other imaging run. This procedure ensures that if two patterns are only differ due to noise, then the expected estimate of the distance is zero (Diedrichsen et al., 2021). For visualization purposes only, the RDM values were square-root transformed and then normalized by their Euclidian norm to remove overall scale differences across ROIs.

Lateralization index. To assess hemispheric asymmetry during syllable production, we computed a lateralization index for each speech-related ROI. Lateralization index was defined as the normalized difference between right- and left-hemisphere activity (or representational

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

distances) within each ROI. To account for potential negative values, the difference was normalized by the sum of the absolute values of both hemispheres:

$$Lateralization\ index = \frac{R - L}{|R| + |L|}$$

Positive values indicate right-hemisphere dominance, while negative values indicate left-hemisphere dominance. Statistical significance of the lateralization was assessed using one-sample t-tests, testing whether the mean lateralization index differed from zero.

Comparing representational dissimilarities. To investigate whether the representational structure of syllables differed between hemispheres we calculated the cosine similarity between each subject's RDM and the leave-one-out group-average RDMs of the left and right hemispheres. This resulted in four similarity measures per subject (per ROI), reflecting how well each hemisphere's representational structure matched the ipsilateral and contralateral group patterns. Differences between ipsilateral and contralateral similarities were assessed using a paired t-test.

In a complementary analysis, we tested whether the superior and inferior cerebellar ROIs differed in their representational structure. For each participant, we computed the cosine similarity between their RDMs and the corresponding leave-one-out group-average RDMs, and compared these similarities using a two-sided paired t-tests.

Model comparison. To assess the contribution of each theoretical model (PoA and VOT) in explaining cortical RDMs, we first strung out the upper triangular elements of the data RDM and the model RDMs into vectors. For each subject and region, we quantified the strength of each feature representation by computing Pearson's correlation between the vectorized empirical RDM and each model RDM. Group level significance in each region was assessed by testing whether the correlation coefficients differed from zero using a one-sided one-sample t-test. To evaluate the relative contribution of each model we performed a non-negative linear regression for each subject, predicting the 15 inter-syllable distances from a linear combination of the two models, including an intercept term to account for baseline similarity:

$$RDM_{region} \sim w_1 + w_2 * PoA + w_3 * VOT$$

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

For this analysis the model RDM vectors (PoA and VOT) were standardized, so that each had a sum of squares equal to 1. The relative contributions of PoA and VOT were then compared using a paired t-test on the regression weights w_2 and w_3 across subjects.

Multidimensional scaling. To visualize the geometric relationships of syllable representations across speech-related regions, we applied classical multidimensional scaling (MDS) to the vectorized RDM vectors. MDS projects the N-dimensional dissimilarity matrix into a lower dimensional space, while preserving pairwise distances. To focus on the pattern of geometric organization rather than overall magnitude of dissimilarities, each RDM was normalized by its total quadric norm (i.e., the square root of the sum of squared Crossnobis distances) before MDS. Dissimilarities were kept as squared Crossnobis distances to ensure unbiased estimates of the neural pattern differences.

Cerebellum-cortex representational similarity. To quantify the similarity between cerebellar and cortical representations, we computed the cosine similarity between each subject's cerebellar RDM and each cortical RDM. To assess whether these similarities exceeded what would be expected under a uniform representational structure, we compared each cerebellar-cortical similarity to a null model in which all pairwise syllable distances were equal. Differences between cerebellar-cortical and null similarities were evaluated using a one-sided paired t-test.

Leave-one-out non-negative regression. To identify which cortical areas contribute unique, non-redundant information in explaining cerebellar RDM vectors, we performed a stepwise non-negative linear regression, predicting cerebellar RDMs from the cortical RDMs, by adding one cortical region at each step. The model was trained using data from all but one subject and tested on the left-out subject. Model performance was evaluated using cosine similarity between the predicted and the observed cerebellar RDMs. To assess whether adding a cortical region significantly improved prediction accuracy, we conducted one-sided paired t-test comparing the performance between consecutive steps.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

Identifying regions responsive to syllable production

We identified six neocortical regions and two cerebellar regions activated during overt syllable repetition (Fig. 2), corresponding to the "minimal speech production network" described by Bohland & Guenther (2006). These areas included a large region in the ventral primary motor (M1) and sensorimotor (S1) cortex (vSM). A second, smaller region was identified more dorsally (dSM), situated between the hand and foot representation. We also observed consistent activity in medial region encompassing parts of the supplementary and pre-supplementary motor areas (SMA), as well as in the frontal and parietal operculum (OP), the sylvian parietal-temporal area (Spt), and the auditory belt (Aud, Fig. 2A). Within the cerebellum, we identified two distinct regions associated with syllable production: a superior region in lobules V/VI (cbSUP) and an inferior region in lobule VIII (cbINF) (Fig. 2B). This dual representation aligns with previous fMRI findings demonstrating two somatomotor maps for tongue movement in the cerebellum (Nettekoven et al., 2024; Saadon-Grosman et al., 2022).

To test which of the regions within this network show hemispheric asymmetry, we calculate a lateralization index for each region separately (Fig. 2C). We found left hemispheric dominance in the SMA ($t_{(11)}$ =-2.63, p=.023) and vSM ($t_{(11)}$ =-2.97, p=.012). Right hemisphere dominance was observed in the operculum ($t_{(11)}$ =2.90, p=.014). In the cerebellum, both superior and inferior regions showed significantly stronger activity in the right hemisphere (cbSUP: p=.045, cbINF: p=.005).

Pattern analysis shows encoding of different syllables

We then asked whether the identified speech-related regions exhibit distinct activity patterns for different syllables. Surface representations of syllable-related activity patterns showed no clear spatial segregation between syllables (Fig. 3A). Instead, the maps revealed individual differences in both the extent and internal organization of activity patches, consistent with previous fMRI findings (Carey et al., 2017).

To assess whether different syllables were reliably represented despite the absence of clear spatial segregation, we quantified syllable-specific reliability across runs and subjects using variance decomposition (Fig. 3B). Importantly, reliability was estimated after subtracting the

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

303 mean activity from each voxel, ensuring the measure reflected the consistency of differences
304 between syllables rather than overall activation levels. All ROIs showed significant positive
305 pattern reliability (see *Methods*, all p<0.05, FDR-corrected), indicating that syllable identity
306 could be reliably decoded from these regions in individual subjects.

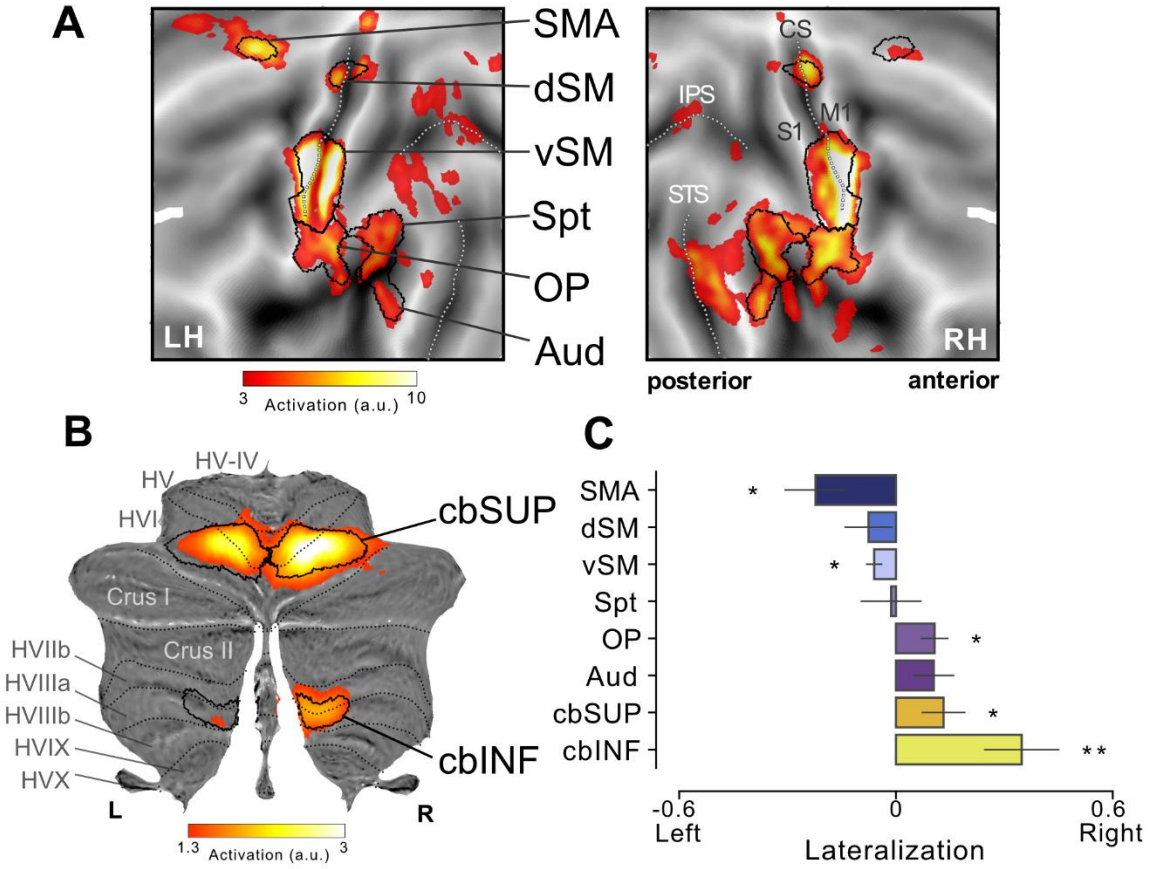

Figure 2. **Average activity during overt syllable production**. Group-average activation maps across all syllable types, projected onto flattened representation of the neocortex (***A***) and the cerebellum (***B***). Major neocortical sulci and cerebellar lobular boundaries are indicated by dotted lines. Boundaries of symmetrically-defined functional ROIs are outlined in black. ***C***, Lateralization index of average activity. Error bar corresponds to standard error of the mean. *p<.05, **p<.01.

Abbreviations: SMA- supplementary motor area; dSM – dorsal sensorimotor; vSM – ventral sensorimotor; OP – operculum; Spt – Sylvian parietal-temporal; Aud – Auditory cortex; CS – central sulcus; S1 – primary sensory cortex; M1 – primary motor cortex; IPS – intraparietal sulcus; STS- superior temporal sulcus; cbSUP – superior cerebellum; cbINF – inferior cerebellum.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We then decomposed pattern reliability into two components: group variance, reflecting structure shared across individuals, and subject-specific variance, capturing reliable but idiosyncratic patterns. In vSM, the group component was highly significant (left: $t_{(11)}$=6.79, $p_{(FDR)}$=.0001; right: $t_{(11)}$=5.93, $p_{(FDR)}$=.0001), accounting for 35.93% (±5.19%) of the individual pattern reliability. Thus, despite the apparent lack of common organization on visual inspection (Fig. 3A), vSM exhibits a systematic topology shared across subjects. To illustrate this, we plotted group contrasts for each place of articulation (bilabial, alveolar, velar) against the remaining syllables. The resulting map (Fig. 3C) shows more dorsal activation for bilabials, and more ventral activity for velars, with alveolar syllables preferentially engaging voxels in between, consistent with previous findings (Bouchard et al., 2013; Carey et al., 2017; Correia et al., 2020; Eichert et al., 2020).

Lastly, we tested whether syllable-specific information differed across hemispheres (Fig. 3D). Significant left lateralization was observed in both SMA and vSM, with stronger information in the left hemisphere compared to the right (SMA: $t_{(11)}$=-2.97, p=.012; vSM: $t_{(11)}$ =-4.15, p=.001). The dSM showed a non-significant trend toward left lateralization ($t_{(11)}$ =-1.66, p=.12). In contrast, the inferior cerebellum exhibited a significant right lateralization, with higher information encoded in the right cerebellar hemisphere ($t_{(11)}$=2.55, p=.026). The operculum, auditory cortex, Spt, and superior cerebellum showed no hemispheric differences in the strength of syllable-specific information (all ps >.1).

Representational geometry in cortical speech regions

Having established reliable syllable-specific information, we next examined its organization across regions using representational similarity analysis (Kriegeskorte & Diedrichsen, 2019). For each ROI, cross-validated Mahalanobis distances between activation patterns for different syllables were computed, yielding a representational dissimilarity matrix (RDM) per region (Fig. 4A). As no significant hemispheric differences were found (see *Methods*, all ps > 0.1), data from left and right hemispheres were averaged to produce a single representational estimate per ROI. To complement this multivariate approach, we also quantified mean activation responses for each syllable within each ROI (Fig. 4B).

We then examined how syllables are represented within each region by comparing the data RDM to two models– place of articulation (PoA) or voice onset time (VOT) (Fig. 4C). First, we

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

338 assessed whether each feature was reflected in the regional representational geometry by
339 correlating the data RDM with each model RDM separately (Fig. 4D). This analysis revealed a
340 strong correspondence with the PoA model in the vSM (mean r=0.67, $p=1.57\times10^{-7}$), the
341 operculum (mean r=0.29, p=.001), and to a lesser degree in the auditory cortex (mean r=0.18,

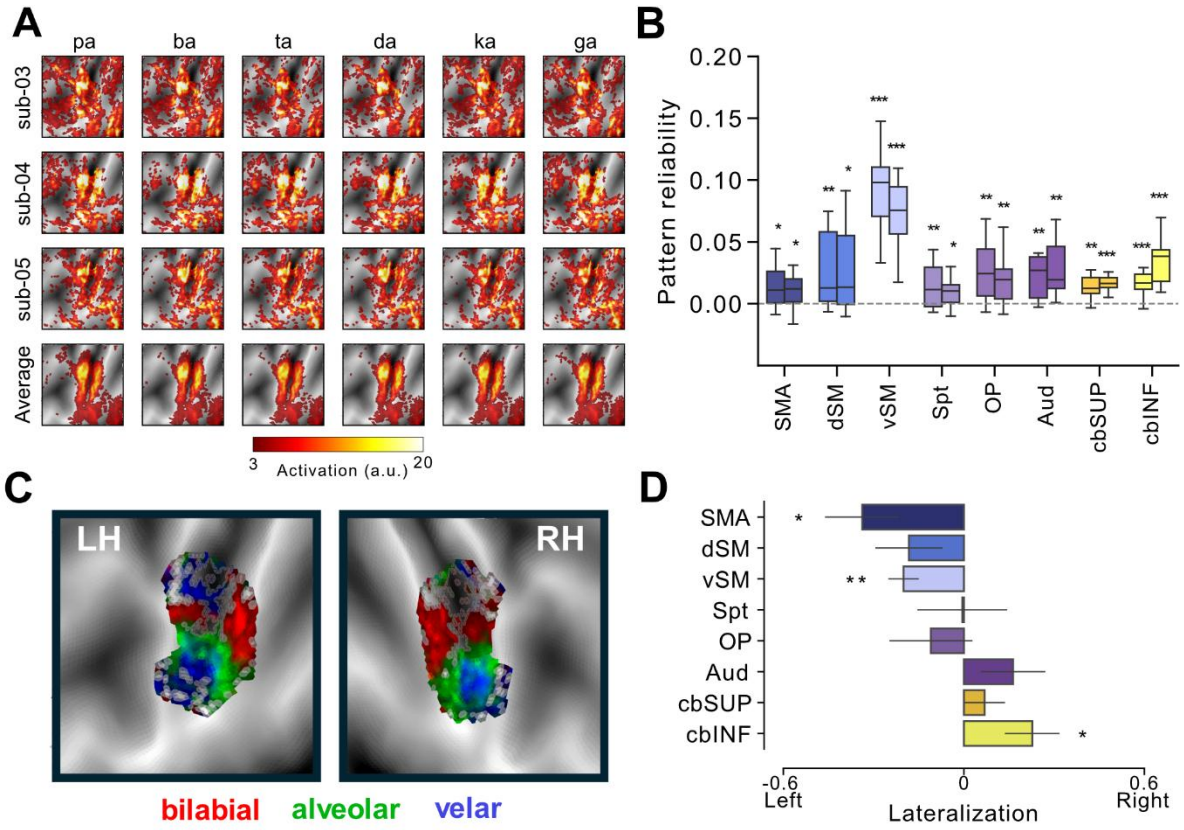

Figure 3. **Evoked activity patterns during the production of different syllables.** ***A***, Activation maps in the left vSM are shown on a flattened neocortical surface. Each row displays activity patterns from an individual participant, with the bottom row showing the group-average across all subjects (N=12). ***B***, Reliability of activity patterns for single syllables in speech ROIs. Pattern reliability reflects the variance explained by shared variance across subjects and task conditions within each subject, across runs. Within each ROI, reliability for the left hemisphere (left boxplot) and right hemisphere (right boxplot) are presented. ***C,*** Group t-maps in left and right vSM showing articulator-selective contrasts for bilabial (red), alveolar (green), and velar (blue). ***D***, Lateralization index calculated as the normalized difference between right and left average Mahalanobis distances, in each ROI. Error bars indicate standard error of the mean. *p<.05, **p<.01, ***p<.001.

Abbreviations: SMA- supplementary motor area; dSM – dorsal sensorimotor; vSM – ventral sensorimotor; OP – operculum; Spt – Sylvian parietal-temporal; Aud – Auditory cortex; cbSUP – superior cerebellum; cbINF – inferior cerebellum; LH – left hemisphere; RH – right hemisphere.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

p=.018), indicating sensitivity to articulatory structure in those regions. Consistent with these representational effects, univariate analysis of mean activity showed significant main effects of PoA in vSM and auditory cortex (Table 1). Post-hoc comparisons revealed distinct response profiles across regions: in the vSM and operculum, alveolar syllables elicited the highest activation, whereas in the auditory cortex, bilabial syllables evoked the lowest activation, and velar syllables the strongest responses (Fig. 4B).

In contrast to PoA representations, significant correlations with VOT were observed in SMA (mean r=0.155, p=.023) and Spt (mean r=0.157, p=.023). The dSM did not show significant correlations with either model, though it was positively correlated with VOT (mean r = 0.15, p=.129). Univariate analyses revealed a significant main effect of VOT in Spt, SMA and dSM, with post-hoc tests showing higher mean activation for voiceless compared to voiced consonants (Table 1, Fig. 4B). Notably, the operculum showed a significant univariate effect for VOT alongside PoA sensitivity in the representational analysis, suggesting an integrated representation of both articulatory place and temporal voicing information in this region. In the auditory cortex, voiced consonants elicited higher activity than voiceless consonants, although this effect did not reach significance.

Next, we evaluated the unique contribution of each feature in explaining the regional RDMs (Fig. 4E). To do this, we used non-negative linear regression to simultaneously fit each cortical RDM with both model RDMs, allowing the two features to compete for shared variance. We then compared the resulting beta weights within each region. vSM showed a significantly stronger weight for PoA compared to VOT ($t_{(11)}$=4.85, p=.0005). The operculum also exhibited a stronger weight to PoA over VOT, albeit to a lesser extent ($t_{(11)}$=2.28, p=.043). Other cortical regions showed no significant difference between PoA and VOT weights (all ps>.1), though the auditory cortex, SMA and dSM showed trends toward VOT dominance.

Multidimensional scaling (MDS; Fig. 4F) summarizes these findings by visualizing the organizational structure of cortical regions in two dimensions: The first dimension contrasts vSM and auditory cortex along the PoA-VOT axis, while the second dimension separates vSM, operculum, and auditory cortex from Spt, SMA and dSM, reflecting a broader distinction between areas sensitive to articulatory versus temporal features.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

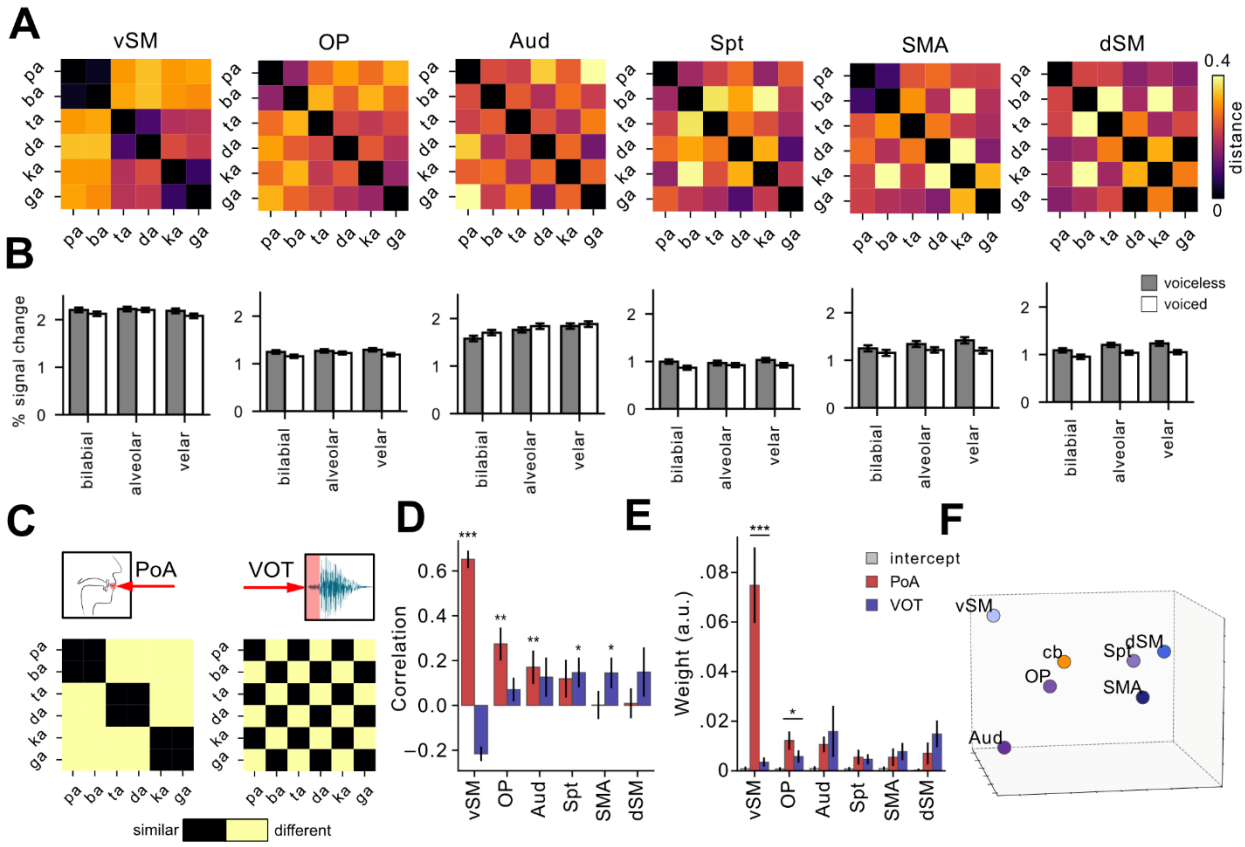

Figure 4. **Representational geometry of syllables within the neocortex**. ***A,*** RDMs between activity patterns evoked by different syllables, averaged across hemispheres and subjects within each region. ***B***, Percent signal change (mean ± SEM) for each region depicted in A, grouped by POA, with voiced (white) and voiceless (gray) consonants shown separately. ***C***, Distance structure for the PoA (left) and VOT (right) models. Black indicated no difference between syllables; yellow indicates large differences. **D,** Pearson's correlation coefficient (mean ± SEM) between each regional RDM and the feature models. PoA correlations are shown in red and VOT correlations in blue. ***E***, Comparison of syllable model fits across speech related ROIs. Bars show mean weight (w ± SEM) across participants for intercept (gray), PoA (red), and VOT (blue). ***F,*** Multidimensional scaling of differences between ROIs in three-dimensional space. *p<.05 ,**p<.01, ***p<.001.

Abbreviations: RDM – representational dissimilarity matrix; vSM – ventral sensorimotor; OP –operculum; Aud – Auditory cortex; Spt – Sylvian parietal-temporal; SMA- supplementary motor area; dSM – dorsal sensorimotor; cb- cerebellum; PoA – place of articulation; VOT – voice onset time.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1. Repeated-measures ANOVA results for place of articulation (PoA) and voicing (VOT) effects on average activity in each region.

Region	Factor	F(df1,df2) [a]	p-value	Post-hoc comparison	Post-hoc p-value[b]
vSM	PoA	F(2,22) = 6.50	**.0060**	Alveolar > Velar	.020
	VOT	F(1,11) = 4.73	.052		
	PoA × VOT	F(2,22) = 1.00	.383		
OP	PoA	F(2,22) = 1.96	.165		
	VOT	F(1,11) = 8.72	**.013**	Voiceless > Voiced	.013
	PoA × VOT	F(2,22) = 0.66	.528		
Aud	PoA	F(2,22) = 24.86	**< .001**	Bilabial < Alveolar Bilabial < Velar	.001 <.001
	VOT	F(1,11) = 1.92	.193		
	PoA × VOT	F(2,22) = 0.44	.650		
Spt	PoA	F(2,22) = 1.11	.348		
	VOT	F(1,11) = 9.75	**.0097**	Voiceless > Voiced	.0097
	PoA × VOT	F(2,22) = 0.75	.485		
SMA	PoA	F(2,22) = 4.91	**.017**	Bilabial > Velar	.011
	VOT	F(1,11) = 12.32	**.0049**	Voiceless > Voiced	.0049
	PoA × VOT	F(2,22) = 1.55	.235		
dSM	PoA	F(2,22) = 6.36	**.0066**	Bilabial > Velar	.025
	VOT	F(1,11) = 17.29	**.0016**	Voiceless > Voiced	.0016
	PoA × VOT	F(2,22) = 0.19	.832		
cb	PoA	F(2,22) = 10.35	**< .001**	Bilabial < Alveolar Bilabial < Velar	.000367 .020
	VOT	F(1,11) = 3.73	.079		
	PoA × VOT	F(2,22) = 0.89	.423		

[a]F values are reported with numerator and denominator degrees of freedom.

[b]Post hoc p values are FDR-corrected.

Representational geometry in the cerebellum

We next investigated syllable representations within the cerebellum. Initially, we compared the superior and inferior cerebellar speech regions and found no significant differences between them ($p>0.3$; Fig. 5A). This aligns with previous studies reporting dual representations of body parts in sensorimotor cerebellar regions, without clear differences in tuning or representational

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

patterns (Nettekoven et al., 2024; Wiestler et al., 2011). Accordingly, we averaged the RDMs across these two cerebellar regions to obtain a more reliable, consolidated measure of syllable organization within the cerebellum (Fig. 5B). Mean activations for each condition were also averaged across these regions to complement the representational analysis with univariate measures (Fig. 5C).

To characterize the geometrical structure of cerebellar syllable representations, we tested their relationships with the feature models. Pearson's correlation revealed a significant correlation between cerebellar RDMs and the PoA model (mean r=0.34, p=.002), but not with the VOT model (mean r=-0.008, p=.537) (Fig. 5D). Univariate analysis supported these findings, showing a main effect of PoA, with lower activation for bilabial sounds (Table 1). Finally, non-negative linear regression demonstrated a significantly stronger weight for PoA compared to VOT ($t_{(11)}$=2.55, p=.026) (Fig. 5E).

The arrangement of representations in speech regions (Fig. 4B) suggests that the representational structure in cerebellar speech areas closely resembles that of the operculum. To quantify these relationships, we calculated cosine similarity between the cerebellum and each cortical region. Because cerebellar syllable-related activity patterns were more variable across subjects than cortical patterns (linear mixed-effects model: β=-0.012, SE=0.006, z=-2.27, p=.023), subsequent analyses were performed at the individual-subject level to capture subject-specific relationships between the cortex and the cerebellum. Within each subject, the cerebellar RDM was significantly more similar to the operculum ($t_{(11)}$=2.206, p=.025) than to a reference model predicting uniform dissimilarities (Fig. 5F). Cerebellar RDMs also showed some similarity to vSM, though this did not reach statistical significance (p=.059). Similarities to other regions were not significant, indicating a selective engagement of the cerebellum with the operculum and vSM during speech production.

Can the cerebellar representation be best described as a mixture of several cortical regions, or is a single region suffice? To address this, we estimated weights that best predicted cerebellar RDMs based on cortical RDMs, using a leave-one-out cross-validation approach incrementally adding ROIs to the model based on their similarity. We found that adding the operculum to a model already including vSM significantly improved prediction of cerebellar RDMs ($t_{(11)}$=2.65, p=.002), indicating that the operculum contributes unique information over and above vSM. In contrast, adding vSM to a model that already included the operculum did not significantly

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

409 improve model fit (p=0.1), suggesting that the operculum could account for most of the syllable
410 representation in the cerebellum (Fig. 5F).
411

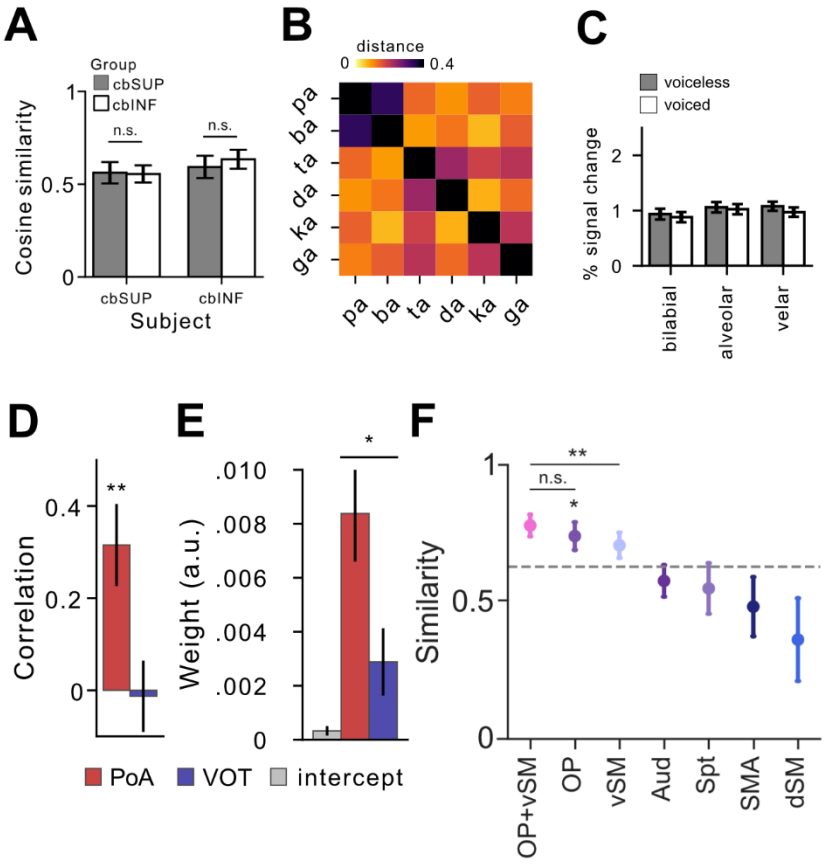

Figure 5. **Representational geometry of syllables within the cerebellum**. ***A,*** Cosine similarities between cerebellar ROIs. Plotted are the cosine similarities of each participant's cbSUP (left) and cbINF (right) RDMs compared to the leave-one-out group RDMs of cbSUP (gray) and cbINF (white). ***B***, RDMs between activity patterns evoked by different syllables, averaged across cbSUP and cbINF. ***C***, Percent signal change (mean ± SEM), grouped by POA, with voiced (white) and voiceless (gray) consonants shown separately. ***D,*** Pearson's correlation coefficient (mean ± SEM) between cerebellar RDM and the feature models. ***E***, Size of model weights (w ± SEM) across participants for intercept (gray), PoA (red), and VOT (blue) in the combined model. ***F***, Cosine similarity between the cerebellum and neocortex. Similarities were calculated between observed and predicted cerebellar RDMs using cross validated stepwise non-negative regression. The dashed gray line represents the similarity of cerebellar RDMs to a null reference, averaged across participants. *p<.05 ,**p<.01, ***p<.001. Abbreviations: RDM – representational dissimilarity matrix; cbINF- inferior cerebellum; cbSUP- superior cerebellum; PoA – place of articulation; VOT – voice onset time; vSM – ventral sensorimotor; OP –operculum; Aud – Auditory cortex; Spt – Sylvian parietal-temporal; SMA- supplementary motor area; dSM – dorsal sensorimotor.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

This study characterized the neural representations of speech during overt syllables production using multivariate analysis and 7T fMRI. We demonstrate that the patterns associated with each syllable showed reliable differences in both cortical and cerebellar speech related regions. These regions exhibited differential tuning along the PoA-VOT axis: While vSM was primarily tuned to place of articulation, dSM, SMA, and Spt were more sensitive to voice onset time. The operculum and auditory cortex demonstrated a hybrid profile, with sensitivity to both features. Furthermore, we found that in the cerebellum, syllables are better represented by their place of articulation rather than voice onset timing. Surprisingly, cerebellar representations aligned most closely with the operculum rather than primary sensorimotor cortex, suggesting that speech control relies on cerebellar-cortical circuits beyond primary motor cortex.

In this study, we found a functional contrast between vSM and dSM in their tuning to speech features. The tuning of the vSM for different places of articulation supports its established role in the motoric sculpting of the vocal tract (Bouchard et al., 2013). Therefore, vSM may function as a motor map with topographical regularities across individuals. The dSM, in contrast, showed greater sensitivity to voice onset time compared to place of articulation (Fig. 4E). Although this effect did not reach statistical significance ($p=.146$), the pattern is consistent with a functional role of the dSM in phonation (Correia et al., 2020). Unlike the somatotopic maps of articulators in vSM, the neural patterns for voice onset time may be more idiosyncratic, potentially reflecting individual-specific strategies for timing respiration and articulation.

Traditionally, the dSM has been associated with trunk movements during respiration (Brown et al., 2009; Loucks et al., 2007). Here, by varying the temporal onset of phonation relative to articulation, we demonstrate that dSM activity reflects the precise temporal gating of airflow during speech production. Recently, the role of the dSM has been extended beyond a simple representation of the trunk. For example, Correia et al., (2020) demonstrated that voiced speech recruits the dSM more heavily than whispered speech, even when controlling for lung volume, suggesting a specific involvement in phonatory control. While we did not directly monitor breathing patterns in the current study, our finding that the dSM is tuned to voice onset time, supports the notion that this region is involved in the synchronization of respiration and laryngeal tension. Future studies utilizing simultaneous fMRI and respiratory tracking are needed to distinguish the dSM's role in global respiration vs. its role in syllable-specific temporal gating.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Contrary to our predictions, cerebellar representations did not exclusively resemble those in vSM, nor did they reflect a mixture of ventral and dorsal sensorimotor representations. Instead, cerebellar representational structure aligned most closely with that of the operculum (Fig. 5F). BOLD signals in the cerebellum predominantly reflect afferent input rather than local output computations (Caesar et al., 2003; Thomsen et al., 2004, 2009). Therefore, the close alignment between opercular and cerebellar representational geometry observed in our data likely reflects information transmitted from the operculum to the cerebellum. Notably, the cerebellum also interacts with the neocortex cortex by sending outputs back to it. However, such efferent interactions are not directly captured by cerebellar BOLD measurements.

According to the DIVA model, opercular-cerebellar interactions support the generation and refinement of feedforward speech motor commands (Guenther, 2006). Within this framework, the frontal operculum is hypothesized to encode speech-motor plans, while the cerebellum contributes to the tuning of these plans. If this account is correct, both regions should be active during speech execution, but only the operculum should show activation during speech preparation. Alternatively, if the cerebellum forms a closed-loop circuit with the operculum, preparatory activation should be observed in both regions.

Clinical evidence further implicates the cerebellum in speech planning. Apraxia of speech and dysarthria are dissociable speech disorders, with apraxia reflecting impaired motor planning, and dysarthria reflecting deficits in motor execution (Ziegler et al., 2012). While dysarthria is classically associated with direct cerebellar damage, highlighting the cerebellum's established role in speech execution (Ackermann, 2008), recent findings in left-hemisphere stroke patients reveal a link between reduced cerebellar gray matter and greater apraxia severity (Gibson et al., 2025). Notably, gray matter volume in right cerebellar lobules V/VI, identified here as the superior cerebellar speech region, best predicted apraxia severity. In contrast, dysarthria severity showed weaker associations with cerebellar gray matter, underscoring the specificity of this relationship to speech planning deficits. Together with its representational alignment to the operculum, these clinical findings strengthen the interpretation that the cerebellum also contributes to speech planning rather than solely to motor execution.

Speech production relies on interactions between the motor and auditory systems, enabling the coordination of articulatory commands with their auditory consequences (Guenther, 2006; Hickok et al., 2011). Our results show that speech production evokes robust and reliable

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

activations in the auditory cortex, primarily located along the medial and lateral Heschl's gyrus (Fig. 2A). Previous studies have shown that this region responds more strongly to speech than non-speech sounds, and encodes acoustic-phonetic features during speech perception (Binder et al., 2000; Mesgarani et al., 2014). Our data extend these findings by demonstrating that the auditory cortex represents phonatory features not only during speech perception, but also during speech production. Because external auditory feedback was largely masked by scanner noise in the current study, these representations are unlikely to arise solely from the airborne auditory feedback. Instead, our results could indicate that the auditory representations are caused by internally generated predictions derived from an efference copy of the articulatory command (Houde et al., 2002). However, these representations may also reflect auditory input transmitted through bone-conductance (v. Békésy, 1949).

In the current study we used a limited set of syllables, focusing exclusively on plosive consonants combined with a low-back vowel (i.e. /a/). While this choice provided strong control over articulatory and temporal dimensions of speech, it limits the extent to which our results can be generalized to other phonemic categories, different manners of articulation, or to more complex, naturalistic, speech. Both vowel identity and manner of articulation evoke distinct activity patterns within the vSM (Bouchard et al., 2013; Bouchard & Chang, 2014). Importantly, because vowel identity shapes consonant articulation via coarticulation, restricting syllables to a single vowel may limit the range of consonant-related representational structure captured here.

In summary, this study reveals feature-specific representational tuning of articulatory and phonatory features within primary sensorimotor areas, with ventral sensorimotor cortex encoding place of articulation and dorsal regions showing sensitivity to voice onset timing. Secondary speech areas, such as the operculum and auditory cortex, exhibited a hybrid representational profile, reflecting heightened sensitivity to the phonetic differences. Surprisingly, cerebellar representations aligned most closely with those of the operculum, suggesting an opercular-cerebellar circuit involved in speech motor planning prior to execution. It remains to be seen which specific computations are performed within the cerebellum and how they influence neocortical speech area.

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgements

This study is supported by the Canadian Institute of Health Research (CIHR Grant # PJT-191815) to JD, and the Canada First Research Excellence Fund (BrainsCAN) to Western University. SJ is further supported by the Council of Higher Education, Israel.

References

Ackermann, H. (2008). Cerebellar contributions to speech production and speech perception: Psycholinguistic and neurobiological perspectives. *Trends in Neurosciences*, *31*(6), 265–272. https://doi.org/10.1016/j.tins.2008.02.011

Ackermann, H., & Brendel, B. (2016). Neurobiology of Speech Production: A Motor Control Perspective. In *Neurobiology of Language* (pp. 741–750). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00059-6

Ariani, G., Shahbazi, M., & Diedrichsen, J. (2025). Cortical Areas for Planning Sequences before and during Movement. *Journal of Neuroscience*, *45*(3). https://doi.org/10.1523/jneurosci.1300-24.2024

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*(1), 95–113. https://doi.org/10.1016/j.neuroimage.2007.07.007

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S., Springer, J. A., Kaufman, J. N., & Possing, E. T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex (New York, N.Y.: 1991)*, *10*(5), 512–528. https://doi.org/10.1093/cercor/10.5.512

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, *32*(2), 821–841. https://doi.org/10.1016/j.neuroimage.2006.04.173

Bosch, J. J. van den, Golan, T., Peters, B., Taylor, J., Shahbazi, M., Lin, B., Charest, I., Diedrichsen, J., Kriegeskorte, N., Mur, M., & Schütt, H. H. (2025). A Python Toolbox for Representational Similarity Analysis. *eLife*, *14*. https://doi.org/10.7554/eLife.107828.1

Bouchard, K. E., & Chang, E. F. (2014). Control of Spoken Vowel Acoustics and the Influence of Phonetic Context in Human Speech Sensorimotor Cortex. *Journal of Neuroscience*, *34*(38), 12662–12677. https://doi.org/10.1523/jneurosci.1219-14.2014

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

535 Bouchard, K. E., Mesgarani, N., Johnson, K., & Chang, E. F. (2013). Functional organization of
536 human sensorimotor cortex for speech articulation. *Nature*, *495*(7441), 327–332.
537 https://doi.org/10.1038/nature11911
538 Brown, S., Laird, A. R., Pfordresher, P. Q., Thelen, S. M., Turkeltaub, P., & Liotti, M. (2009).
539 The somatotopy of speech: Phonation and articulation in the human motor cortex. *Brain*
540 *and Cognition*, *70*(1), 31–41. https://doi.org/10.1016/j.bandc.2008.12.006
541 Caesar, K., Gold, L., & Lauritzen, M. (2003). Context sensitivity of activity-dependent increases
542 in cerebral blood flow. *Proceedings of the National Academy of Sciences*, *100*(7), 4239–
543 4244. https://doi.org/10.1073/pnas.0635075100
544 Carey, D., Krishnan, S., Callaghan, M. F., Sereno, M. I., & Dick, F. (2017). Functional and
545 Quantitative MRI Mapping of Somatomotor Representations of Human Supralaryngeal
546 Vocal Tract. *Cerebral Cortex*, cercor;bhw393v2. https://doi.org/10.1093/cercor/bhw393
547 Correia, J. M., Caballero-Gaudes, C., Guediche, S., & Carreiras, M. (2020). Phonatory and
548 articulatory representations of speech production in cortical and subcortical fMRI
549 responses. *Scientific Reports*, *10*(1), 4529. https://doi.org/10.1038/s41598-020-61435-y
550 Diedrichsen, J. (2006). A spatially unbiased atlas template of the human cerebellum.
551 *NeuroImage*, *33*(1), 127–138. https://doi.org/10.1016/j.neuroimage.2006.05.056
552 Diedrichsen, J., Berlot, E., Mur, M., Schütt, H. H., Shahbazi, M., & Kriegeskorte, N. (2021).
553 *Comparing representational geometries using whitened unbiased-distance-matrix*
554 *similarity* (arXiv:2007.02789). arXiv. https://doi.org/10.48550/arXiv.2007.02789
555 Diedrichsen, J., & Zotow, E. (2015). Surface-Based Display of Volume-Averaged Cerebellar
556 Imaging Data. *PLOS ONE*, *10*(7), e0133402.
557 https://doi.org/10.1371/journal.pone.0133402
558 Eichert, N., Papp, D., Mars, R. B., & Watkins, K. E. (2020). Mapping Human Laryngeal Motor
559 Cortex during Vocalization. *Cerebral Cortex*, *30*(12), 6254–6269.
560 https://doi.org/10.1093/cercor/bhaa182
561 Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-resolution intersubject
562 averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, *8*(4),
563 272–284. https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<272::AID-
564 HBM10>3.0.CO;2-4

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

565 Gibson, M., Maydeu-Olivares, A., Newman-Norlund, R., & Rorden, C. (2025). Beyond the
566 cortex: Cerebellar contributions to apraxia of speech in stroke survivors. *Cortex*, *191*, 1–
567 11. https://doi.org/10.1016/j.cortex.2025.07.004
568 Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil,
569 K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C.
570 (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–
571 178. https://doi.org/10.1038/nature18933
572 Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds.
573 *Journal of Communication Disorders, ASHA 2005 Research Symposium: Physiological*
574 *Foundations of Speech Motor Development and Production*, *39*(5), 350–365.
575 https://doi.org/10.1016/j.jcomdis.2006.06.013
576 Guenther, F. H., & Hickok, G. (2016). Chapter 58—Neural Models of Motor Speech Control. In
577 G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 725–740). Academic
578 Press. https://doi.org/10.1016/B978-0-12-407794-2.00058-4
579 Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor Integration in Speech Processing:
580 Computational Basis and Neural Organization. *Neuron*, *69*(3), 407–422.
581 https://doi.org/10.1016/j.neuron.2011.01.019
582 Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the
583 Auditory Cortex during Speech: An MEG Study. *Journal of Cognitive Neuroscience*,
584 *14*(8), 1125–1138. https://doi.org/10.1162/089892902760807140
585 Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., & Turner, R. (2002). Image
586 Distortion Correction in fMRI: A Quantitative Evaluation. *NeuroImage*, *16*(1), 217–240.
587 https://doi.org/10.1006/nimg.2001.1054
588 Jürgens, U. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral*
589 *Reviews*, *26*(2), 235–258. https://doi.org/10.1016/s0149-7634(01)00068-9
590 Kelly, R. M., & Strick, P. L. (2003). Cerebellar Loops with Motor Cortex and Prefrontal Cortex
591 of a Nonhuman Primate. *Journal of Neuroscience*, *23*(23), 8432–8444.
592 https://doi.org/10.1523/JNEUROSCI.23-23-08432.2003
593 King, M., Shahshahani, L., Ivry, R. B., & Diedrichsen, J. (2023). A task-general connectivity
594 model reveals variation in convergence of cortical inputs to functional regions of the
595 cerebellum. *eLife*, *12*, e81511. https://doi.org/10.7554/eLife.81511

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the Onion of Brain Representations. *Annual Review of Neuroscience*, *42*, 407–432. https://doi.org/10.1146/annurev-neuro-080317-061906

Loucks, T. M. J., Poletto, C. J., Simonyan, K., Reynolds, C. L., & Ludlow, C. L. (2007). Human brain activation during phonation and exhalation: Common volitional control for two upper airway functions. *NeuroImage*, *36*(1), 131–143. https://doi.org/10.1016/j.neuroimage.2007.01.049

Mesgarani, N., Cheung, C., Johnson, K., & Chang, E. F. (2014). Phonetic Feature Encoding in Human Superior Temporal Gyrus. *Science*, *343*(6174), 1006–1010. https://doi.org/10.1126/science.1245994

Nettekoven, C., Zhi, D., Shahshahani, L., Pinho, A. L., Saadon-Grosman, N., Buckner, R. L., & Diedrichsen, J. (2024). A hierarchical atlas of the human cerebellum for functional precision mapping. *Nature Communications*, *15*(1), 8376. https://doi.org/10.1038/s41467-024-52371-w

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., & Kriegeskorte, N. (2014). A Toolbox for Representational Similarity Analysis. *PLOS Computational Biology*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Penfield, W., & Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, *60*(4), 389–443. https://doi.org/10.1093/brain/60.4.389

Saadon-Grosman, N., Angeli, P. A., DiNicola, L. M., & Buckner, R. L. (2022). A third somatomotor representation in the human cerebellum. *Journal of Neurophysiology*, *128*(4), 1051–1073. https://doi.org/10.1152/jn.00165.2022

Thomsen, K., Offenhauser, N., & Lauritzen, M. (2004). Principal neuron spiking: Neither necessary nor sufficient for cerebral blood flow in rat cerebellum. *The Journal of Physiology*, *560*(1), 181–189. https://doi.org/10.1113/jphysiol.2004.068072

Thomsen, K., Piilgaard, H., Gjedde, A., Bonvento, G., & Lauritzen, M. (2009). Principal Cell Spiking, Postsynaptic Excitation, and Oxygen Consumption in the Rat Cerebellar Cortex. *Journal of Neurophysiology*, *102*(3), 1503–1512. https://doi.org/10.1152/jn.00289.2009

bioRxiv preprint doi: https://doi.org/10.64898/2026.01.30.702863; this version posted January 31, 2026. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

v. Békésy, G. (1949). The Structure of the Middle Ear and the Hearing of One's Own Voice by Bone Conduction. *The Journal of the Acoustical Society of America*, *21*(3), 217–232. https://doi.org/10.1121/1.1906501

Van Essen, D. C., Glasser, M. F., Dierker, D. L., Harwell, J., & Coalson, T. (2012). Parcellations and Hemispheric Asymmetries of Human Cerebral Cortex Analyzed on Surface-Based Atlases. *Cerebral Cortex*, *22*(10), 2241–2262. https://doi.org/10.1093/cercor/bhr291

Wiestler, T., McGonigle, D. J., & Diedrichsen, J. (2011). Integration of sensory and motor representations of single fingers in the human cerebellum. *Journal of Neurophysiology*, *105*(6), 3042–3053. https://doi.org/10.1152/jn.00106.2011

Ziegler, W., Aichert, I., & Staiger, A. (2012). Apraxia of Speech: Concepts and Controversies. *Journal of Speech, Language, and Hearing Research*, *55*(5), S1485–S1501. https://doi.org/10.1044/1092-4388(2012/12-0128)