Differentiating neural substrates related to phonological working memory and motor programming in speech sequence learning

Haochen Wan[1], Jason A. Tourville[2], Jackie S. Kim[2], Barbara Holland[2], Alfonso Nieto-Castañón[2,3], Frank H. Guenther[1,2,4]

[1]Department of Biomedical Engineering, Boston University, Boston, MA; [2]Department of Speech, Language & Hearing Sciences, Boston University, Boston, MA;
[3]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA; [4]Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA

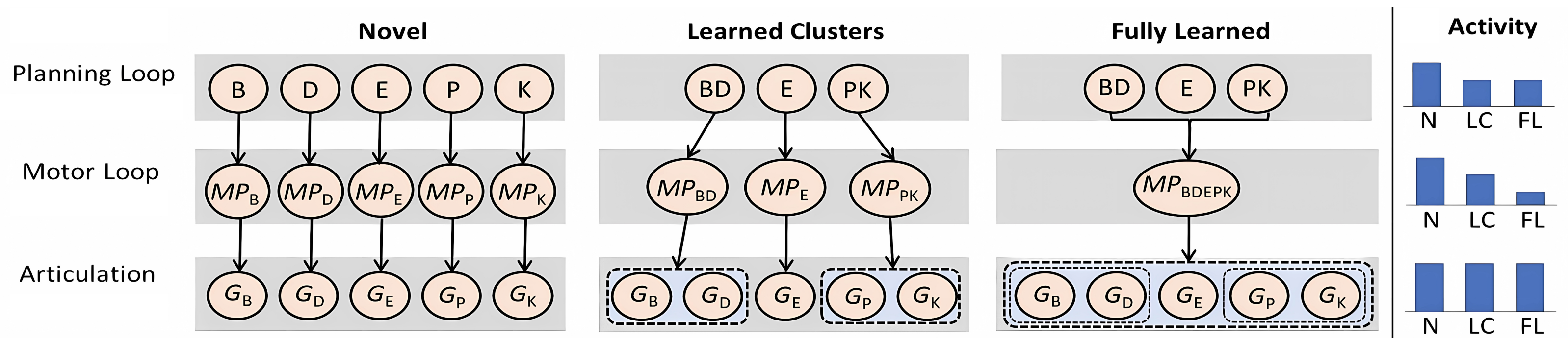
BOSTON UNIVERSITY

Introduction

- To facilitate fluent speech, **frequently produced sequences** are thought to be **encoded as "chunks"** to conserve neural resources, rather than built from scratch each time.[1]
- As speech sequencing encompasses multiple stages[2]—**phonological encoding**, **phonetic encoding**, and **motor execution**—different neural mechanisms and chunk units may be used across these stages.
- Prior behavioral studies from our lab supported an Onset–Nucleus–Coda (**O-N-C**) syllabic frame structure in phonological working memory (**PWM**), and a **syllable-frame-independent representation** as units for **motor programs**.[3,4]
- Integrating previous neuroimaging and behavioral literatures, our lab developed a **computational model named GODIVA**[5], which addresses neural mechanisms underlying **short-term buffering** of upcoming speech sounds in **working memory** as well as **selecting** and timely **initiating** of the corresponding **motor programs**.

- The most important regions of interest (ROIs) posited by the GODIVA model include cortical regions such as *supplementary motor area (**SMA**)*, ***pre-SMA***, *posterior inferior frontal sulcus (**pIFS**)*, *ventral premotor cortex (**vPMC**)*, and subcortical structures including ***caudate***, ***putamen***, *globus pallidus (**GP**)*, *ventral anterior nucleus (**VA**)*, and *ventral lateral nucleus (**VL**)*.
- Previous neuroimaging studies [e.g.,6] **contrasting novel (N) and fully learned (FL)** syllables **couldn't differentiate** areas associated with PWM (planning loop) with those associated with motor programs (motor loop) due to the nature of this contrast, as illustrated in **Fig 1**.
- Here, we introduce a 3rd condition, **learned clusters (LC)**, of which clusters were practiced in a different syllable, aiming to differentiate the neural mechanisms behind the planning and motor stages of the sub-syllabic sequencing process.

Fig 1: Expected BOLD pattern in brain areas along speech sequencing hierarchy

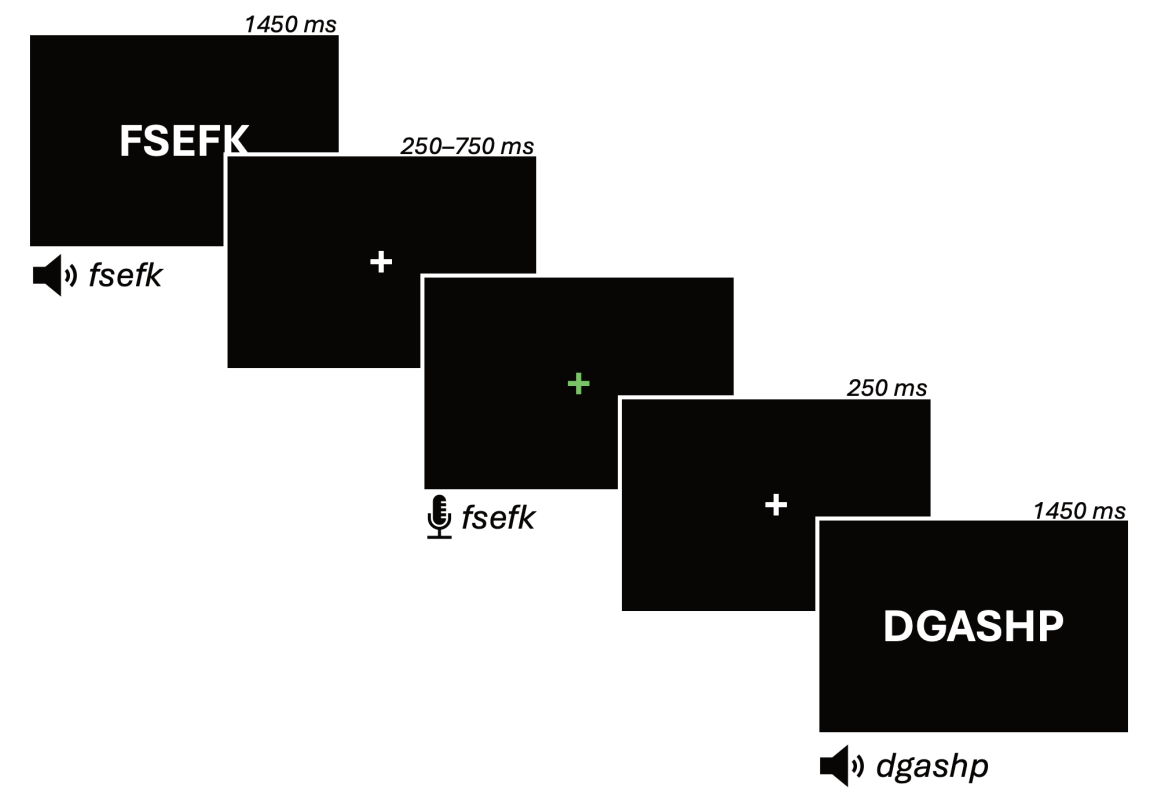


Methods

Participants

- 20 American English native speakers (13F/7M, 21.1±1.6 yrs)
- No prior exposure to languages used to design stimuli
- No speech/hearing disorders or MRI contraindications

Fig 2: Syllable repetition task

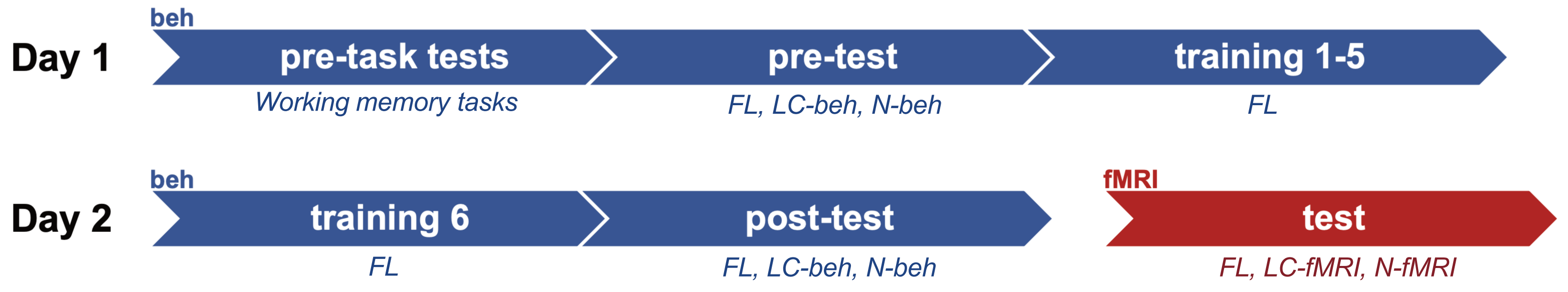


Task details

- **16 CCVCC** (C: consonant, V: vowel) syllables [e.g., **FSEFK**] with different phonotactically illegal consonant clusters (CCs) divided into **4 groups**; each participant trained with 4 syllables (**FL** condition)
- Within each group, recombinations of onset and coda CCs created 32 additional syllables; used as **LC** and **N**
- In total, 146 syllables (36 x 4 + 2 practice) created; model production recorded by a phonetically trained, female American English native speaker
- Stimuli lists counterbalanced among participants

- Each **training** syllable repeated **10** times/run x **6 runs** across two consecutive days (see **Figs 2** and **3**)
- After the post-test, participants completed **5 test sessions** in the MRI, with sparse-sampling acquisition triggered 4.5 s after each speech onset (TR = 1.59 s)
- fMRI test runs used **different N** list from behavioral sessions (pre- and post-tests) to keep them "novel"

Fig 3: Flow of experiment sessions



Day 1

beh

pre-task tests — *Working memory tasks*

pre-test — *FL, LC-beh, N-beh*

training 1-5 — *FL*

Day 2

beh

training 6 — *FL*

post-test — *FL, LC-beh, N-beh*

fMRI

test — *FL, LC-fMRI, N-fMRI*

Results

- Trials were **removed** from all further analyses **if**: (a) Speech started before GO signal; (b) Speech didn't finish by the end; (c) No response during a speech trial; (d) Speech-like sound during a rest trial [only in fMRI test sessions]
- Trials were marked as **inaccurate if**: (a) Overall disfluent [e.g., restart or unrecognizable]; (b) Phoneme omitted; (c) Additional consonant inserted; (d) Phoneme substituted; (e) Phonemes in incorrect order; and (f) Incorrectly voiced/unvoiced [e.g., k↔g]

Fig 4: Training improved performance

Dots are outliers identified using interquartile range

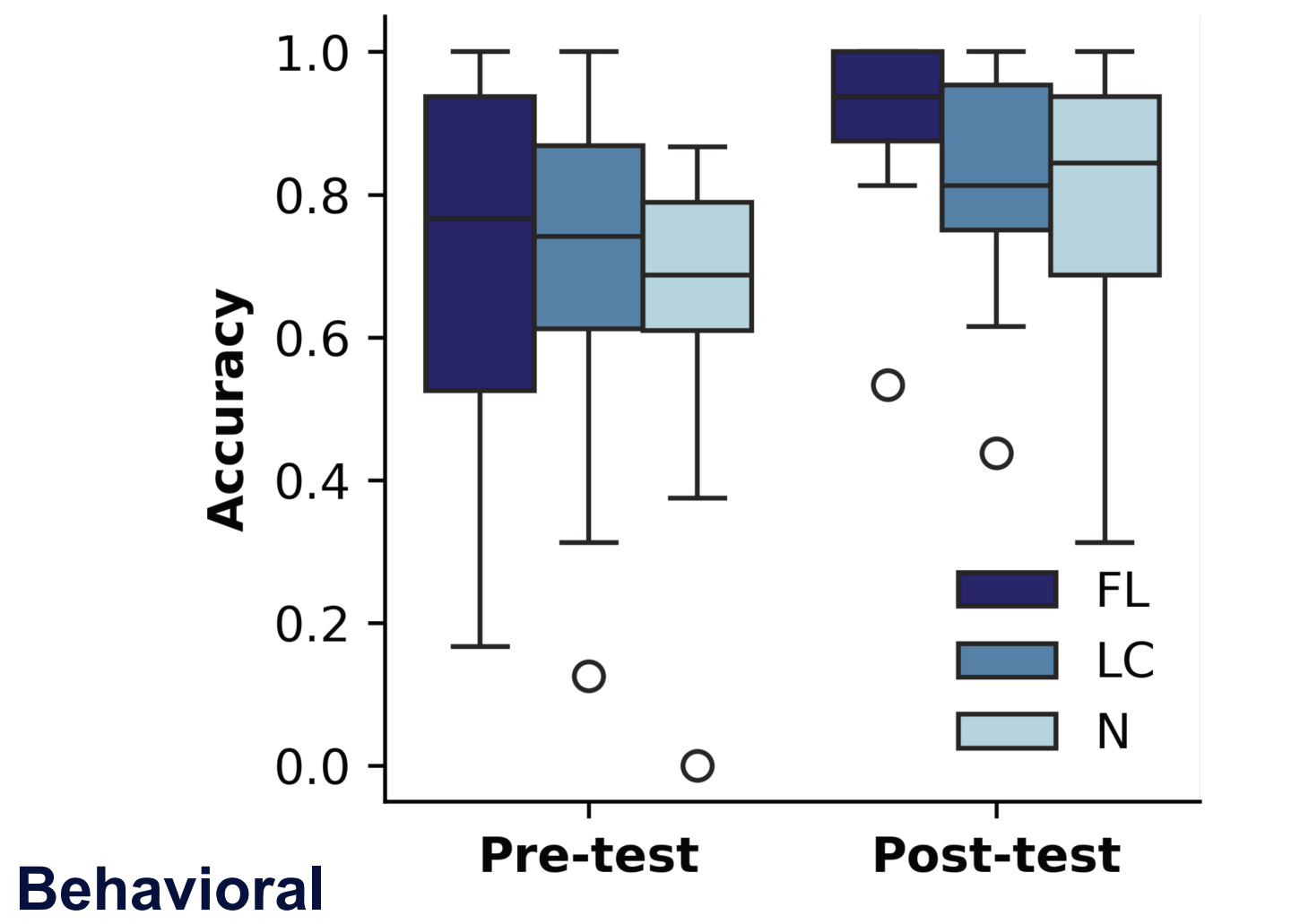
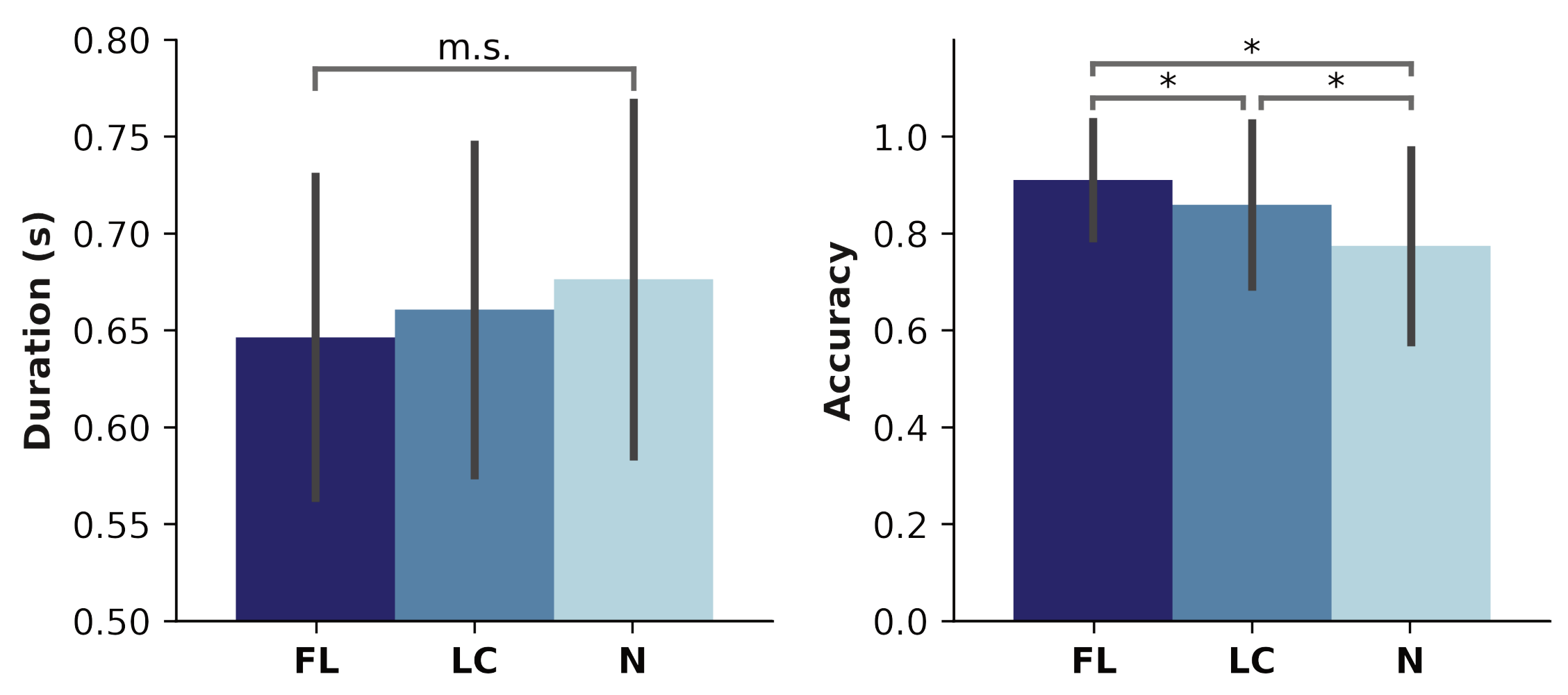


Fig 5: Utterance duration and accuracy during fMRI

Error bars: standard deviation; 5 fMRI sessions pooled



Behavioral

- Training significantly improved accuracy comparing pre-test and post-test (ANOVA; F = 20.89, $p < 0.001$, **Fig 4**)
- For 5 testing sessions, ANOVA on **utterance duration** showed marginal significant effect of conditions (F = 2.89, $p = 0.057$); post-hoc *Fisher's Least Significant Difference* (LSD) yielded significant difference only in the pair of **FL vs. N [Fig 5 left]**
- Similarly, ANOVA on **accuracy** showed significant effect of conditions (F = 16.85, $p < 1e-6$); post-hoc LSD yielded significant difference in all of 3 contrasts: **FL vs. N**, **LC vs. N**, and **FL vs. LC [Fig 5 right]**

Neuroimaging

- Compared to **FL**, production of **N** syllables required higher activity in *vPMC*, *anterior insula (aINS)*, *posterior superior temporal gyrus (pSTG)*, *intraparietal sulcus (IPS)*, and *pre-SMA* in both hemispheres; *pIFS* and *anterior IFS (aIFS)* in left hemisphere [**Fig 6 top**].
- Compared to **LC**, production of **N** syllables required higher activity in *vPMC*, and *pre-SMA* in both hemispheres; *pIFS* and *IPS* mostly in left hemisphere [**Fig 6 center**].
- Compared to **FL**, production of **LC** syllables required higher activity in *vPMC*, *IPS* and *aIFS* only in left hemisphere [**Fig 6 bottom**].
- In addition to whole-brain surface-based analyses, we also conducted ROI-based subcortical analyses based on hypotheses from the GODIVA model[5] and/or prior imaging studies.[6,7] Subcortical results are summarized in **Fig 7**, with significant ones in red ($p < 0.05$, not corrected for multiple ROIs). The definition of several important ROIs are shown in **Fig 8**.

Fig 6: Cortical fMRI results

Permutation-based cluster-level inference, vert p < 0.001, cluster p < 0.05 (FDR-corrected)

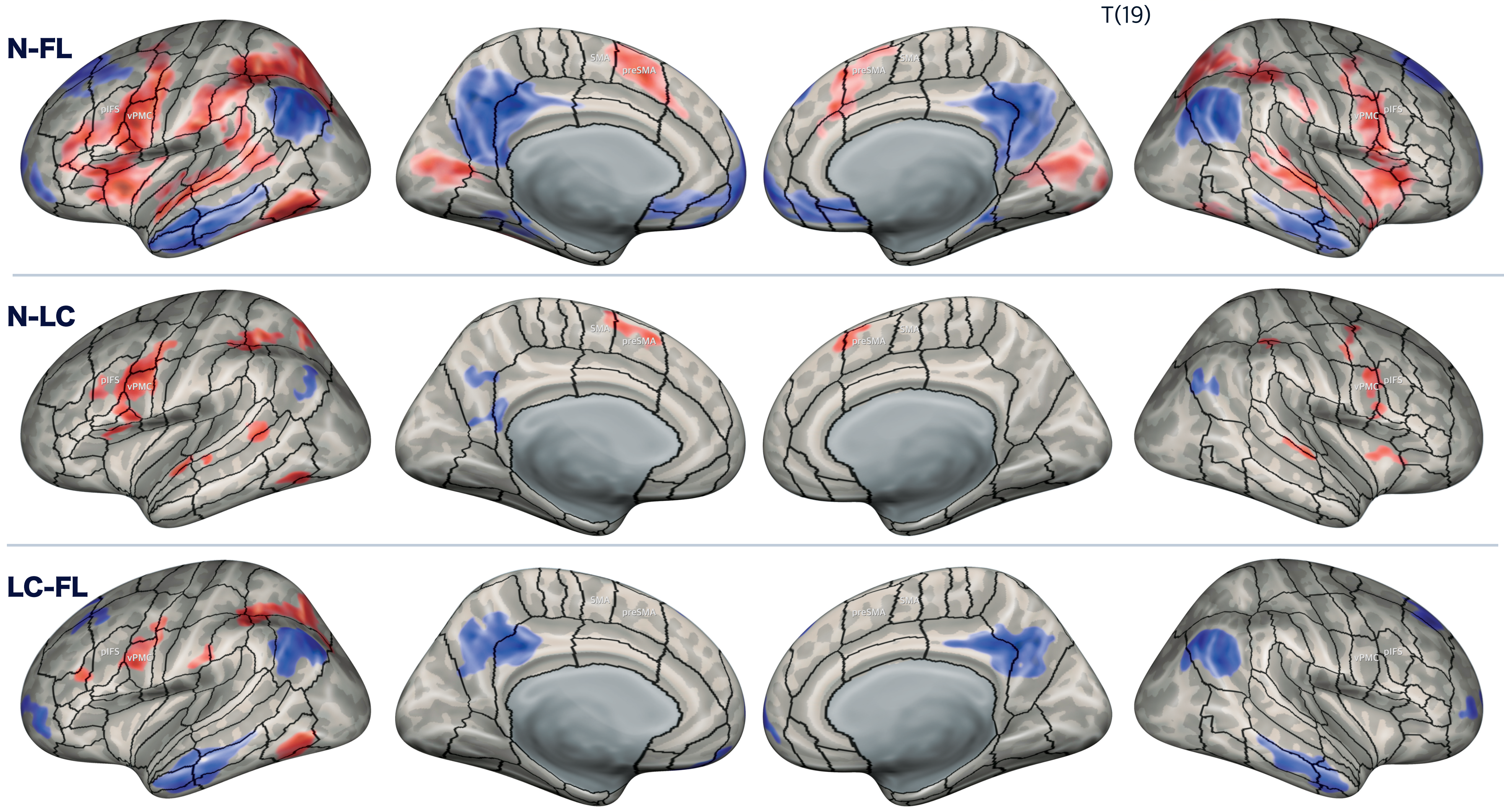


Fig 7: Subcortical fMRI results

Error bars: 95% CI; significantly different ROIs in red

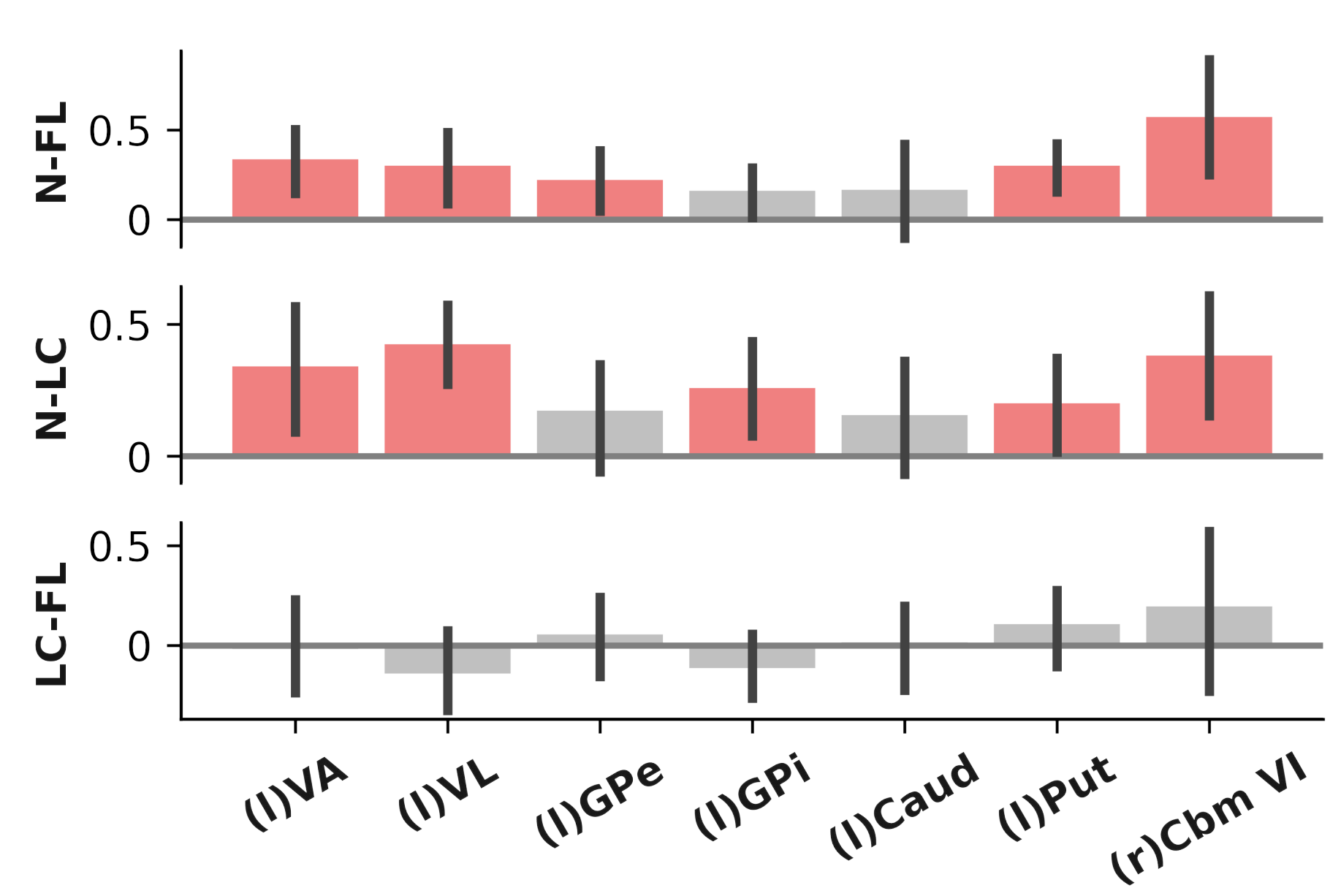
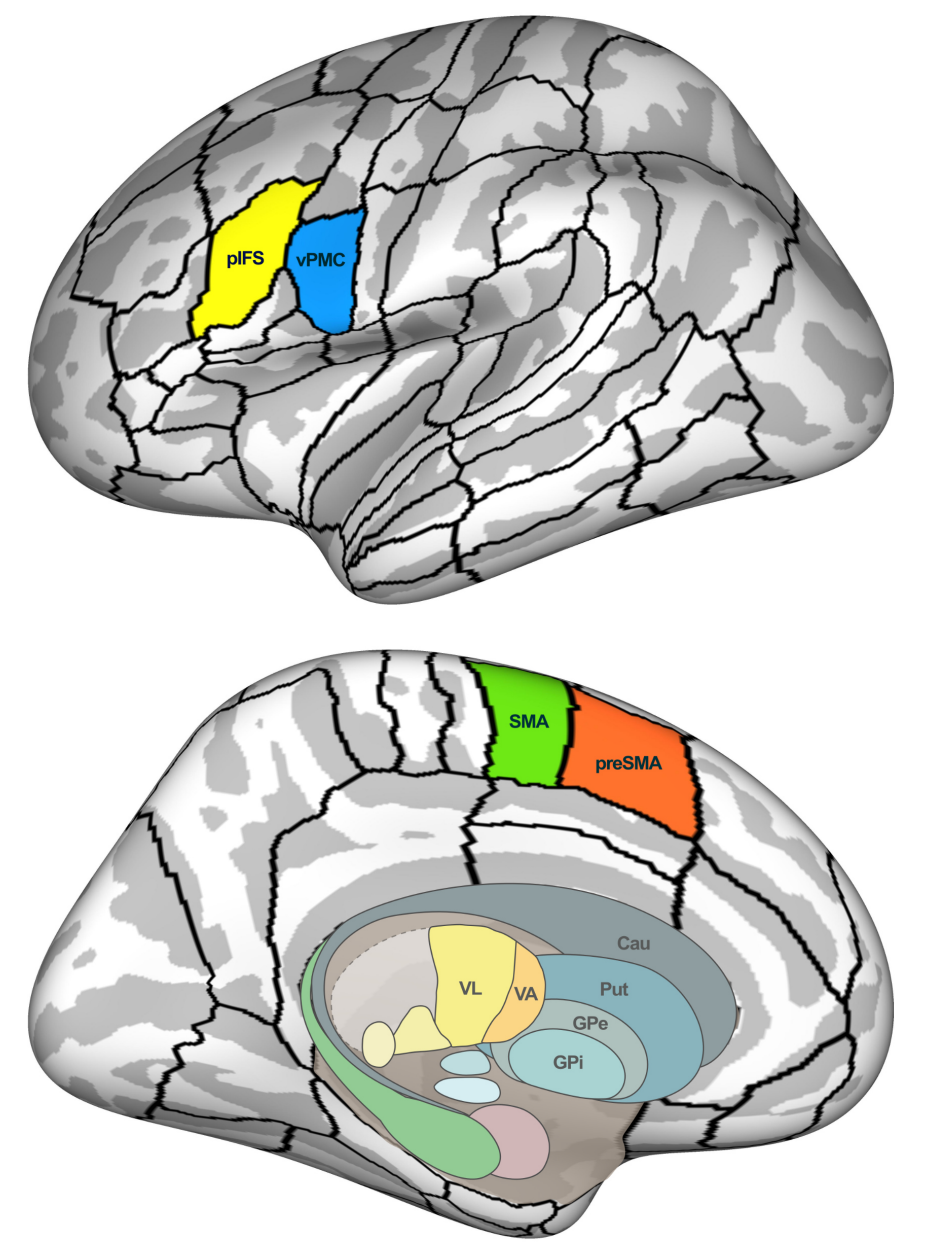


Fig 8: ROI definitions



Discussion

- **Left *pIFS*** activities followed the pattern of N > LC ≈ FL, consistent with its posited function as a **phonological content buffer**[5] (*planning loop*)
- ***pre-SMA*** activities in both hemispheres followed the pattern of N > LC ≈ FL, consistent with its posited function as a **sequential structure buffer**[5] (*planning loop*)
- **Left *vPMC*** activities followed the pattern of N > LC > FL, consistent with its posited function as a **speech sound map** containing motor target as well as auditory and somatosensory targets for production[5] (*motor loop*)
- ***SMA*** was originally proposed to initiate the motor program by activating appropriate nodes in the speech sound map[5], which associated it with the motor loop in GODIVA. However, *SMA* activities here (in both hemispheres) followed the pattern of N ≈ LC ≈ FL, consistent with an alternative hypothesis of it being tightly associated with **single motor acts** rather than well-learned action sequences[8], putting it in the *articulation* level in **Fig 1**.

- Other ROIs that showed significant differences include ***aINS***, of which function in speech sequencing is largely under debate, and ***IPS***, which is linked to working memory and internal prediction model.

- In addition to the expected results that speech related areas are more active for syllables with heavier processing loads (**N** > **LC**, **LC** > **FL**), another interesting finding is that a typical default mode network (**DMN**) composition showed up as more recruited for lighter-load conditions [blue areas in **Fig 8**].

- Future work includes a deeper look into each **subcortical area's involvement** in speech sequence learning and **correlation** analyses between **behavioral** (e.g., accuracy, utterance duration, or their increment due to training) and **neuroimaging** measures (e.g., cortical thickness from sMRI, functional connectivity from rs-fMRI, white matter bundle thickness from dMRI).

References

[1] Guenther, F. H. (2016). Neural Control of Speech. The MIT Press. https://doi.org/10.7551/mitpress/10471.001.0001

[2] Levelt, W. J. M. (1989). Speaking: From intention to articulation (pp. xiv, 566). The MIT Press.

[3] Segawa, J., Masapollo, M., Tong, M., Smith, D. J., & Guenther, F. H. (2019). Chunking of phonological units in speech sequencing. Brain and Language, 195, 104636. https://doi.org/10.1016/j.bandl.2019.05.

[4] Masapollo, M., Zezas, E., Shamsi, A., Wayland, R., Smith, D. J., & Guenther, F. H. (2023). Disentangling Effects of Memory Storage and Inter-articulator Coordination on Generalization in Speech Motor Sequence Learning. Journal of Psycholinguistic Research, 52(6), 2181–2210. https://doi.org/10.1007/s10936-023-09998-5

[5] Bohland, J. W., Bullock, D., & Guenther, F. H. (2010). Neural representations and mechanisms for the performance of simple speech sequences. Journal of Cognitive Neuroscience, 22(7), 1504–1529. https://doi.org/10.1162/jocn.2009.21306

[6] Segawa, J. A., Tourville, J. A., Beal, D. S., & Guenther, F. H. (2015). The Neural Correlates of Speech Motor Sequence Learning. Journal of Cognitive Neuroscience, 27(4), 819–831. https://doi.org/10.1162/jocn_a_00737

[7] Masapollo, M., Segawa, J. A., Beal, D. S., Tourville, J. A., Nieto-Castañón, A., Heyne, M., Frankford, S. A., & Guenther, F. H. (2021). Behavioral and Neural Correlates of Speech Motor Sequence Learning in Stuttering and Neurotypical Speakers: An fMRI Investigation. Neurobiology of Language, 2(1), 106137. https://doi.org/10.1162/nol_a_00027

[8] Tanji, J. (2001). Sequential organization of multiple movements: Involvement of cortical motor areas. Annual Review of Neuroscience, 24, 631–651. https://doi.org/10.1146/annurev.neuro.24.1.631

NIH NIDCD

This work was supported by the NIDCD
of the National Institutes of Health
R01 DC007683 and **T32 DC013017-07**

Guenther Lab

website *e-poster*