Second Biennial Boston Speech Motor Control Symposium, Boston, MA, USA, June 2019

On the nature of working memory structures in speech sequencing

Matthew Masapollo[1], Dante Smith[2], & Frank H. Guenther[1,3]

BOSTON UNIVERSITY

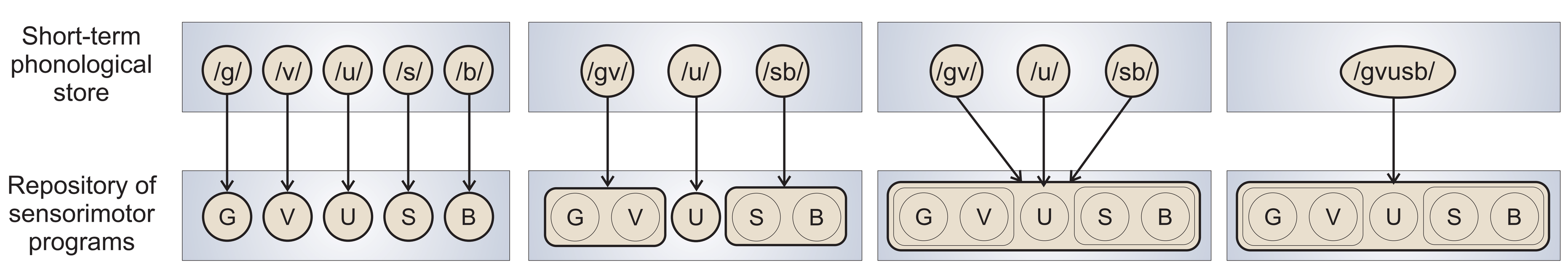
[1]Department of Speech, Language & Hearing Sciences, Boston University
[2]Graduate Program in Neuroscience, Boston University
[3]Department of Biomedical Engineering, Boston University

BACKGROUND

- Theories of language and speech production commonly propose that speakers plan and execute complex or extended sequences of phonemic segments by temporally integrating and consolidating individual segments into cohesive working memory structures or "chunks," which reduces processing load and improves motor performance (e.g., Shattuck-Hufnagel, 1992; Bohland, Bullock, & Guenther, 2010; Guenther, 2016).
- While researchers generally share the assumption that chunk segmentation and concatenation underlie serial speech planning and production, no consensus has yet been reached regarding the mechanisms that support these processes, or the structure and format of the chunks that these processes operate on (see, Segawa, Masapollo, Tong, Smith & Guenther, 2019, for recent discussion).
- **Characterizing these structures will provide critical information, currently lacking, on speech sequencing, thus advancing fundamental knowledge relevant to understanding sequencing deficits in motor speech disorders such as stuttering and AOS, which may be partly due to problems with chunking speech movements** (see, e.g., Buchwald *et al*., 2017).

Research Question:

- **What is the precise nature of the working memory structures involved in speech sequence assembly? i.e., phonemes, clusters, entire syllables?**



- To begin to identify these structures, we investigated generalization of motor learning from training to transfer utterances. During training, speakers repeated *isolated* CCVCC syllables containing non-native consonant clusters. After learning, we tested for generalization under a higher memory load by having speakers repeat four types of *pairs* of syllables that overlapped to varying degrees with the practiced syllables:
  1. *Practiced CCVCC*: syllables practiced in their entirety during training
  2. *Practiced O/C*: novel syllables with practiced clusters that occur in the same syllable position (onset or coda) as in training
  3. *Practiced CC*: novel syllables with practiced clusters but in a different position from training
  4. *Novel CCVCC*: novel syllables with novel clusters

See Results section for the schematized predictions.

METHODS

Stimuli: CCVCC syllables containing non-native onset and coda clusters (e.g., shkizg)

Procedure & Design: Cued sequence learning paradigm (see Segawa *et al*., 2019)

- Speakers intensively practiced producing a set of syllables over two consecutive days (training phase), and then we assessed transfer of learning to the aforementioned types of pairs of syllables involving non-native clusters (test phase).
- On each trial, speakers saw an orthographic display of the syllable(s) in tandem with its corresponding auditory prompt. Subjects only heard each prompt once. After the offset of the auditory presentation, a tone was presented for 50 ms. The onset of the tonal stimulus was randomly jittered between 1,500 and 2,000 ms in the training trials and between 2,500 and 3,000 ms in the test trials. This tone served as a go-signal for the subject to go ahead and repeat the token as clearly and accurately as possible.

Subjects: Native, monolingual American English speaking adults

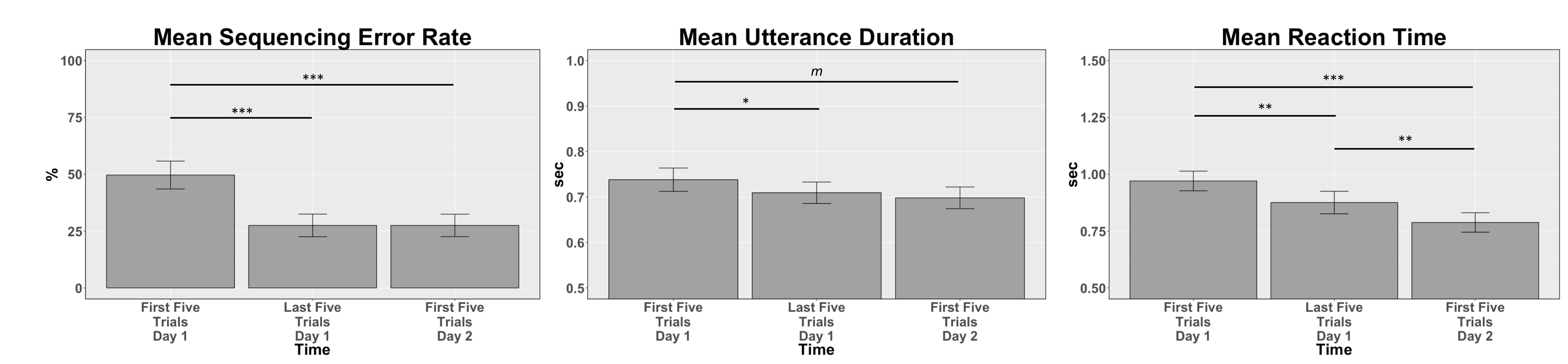
Data Processing: We used Praat and custom MATLAB software to perceptually rate and acoustically measure onsets and offsets of syllables by viewing the waveform and spectrogram and listening to the audio files. All elicited utterances were hand-labelled for:
1. *Sequencing* errors (e.g., phoneme omissions, substitutions, etc.)
2. *Non-sequencing* (i.e., fluency) errors (e.g., vocoid epenthesis)

*Note: Since the primary goal of the current study was to investigate the cognitive structures and processes involved in speech sequencing, we focused our analyses on *sequencing* errors (see Results).

RESULTS

Training Phase

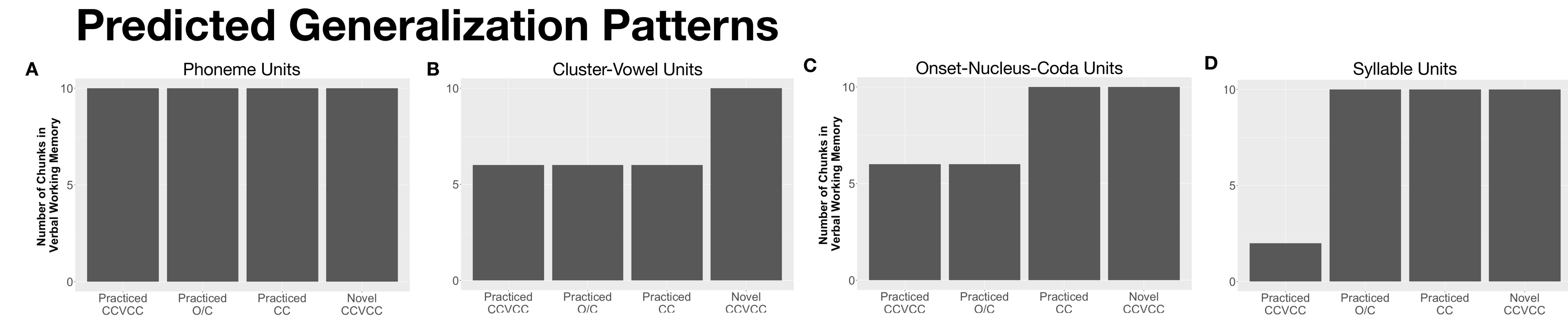
- The dependent variables were ***mean sequencing error rate, utterance duration*** (i.e., time from utterance onset to offset) and ***reaction time*** (i.e., time from the offset of the go-signal to utterance onset); we computed scores for each performance measure by averaging the values obtained (for each syllable) on the first five and last five training trials on day one and the first five training trials on day two. Duration and RT measures were only analyzed for the first or last five utterances coded as *properly sequenced* productions on each day.



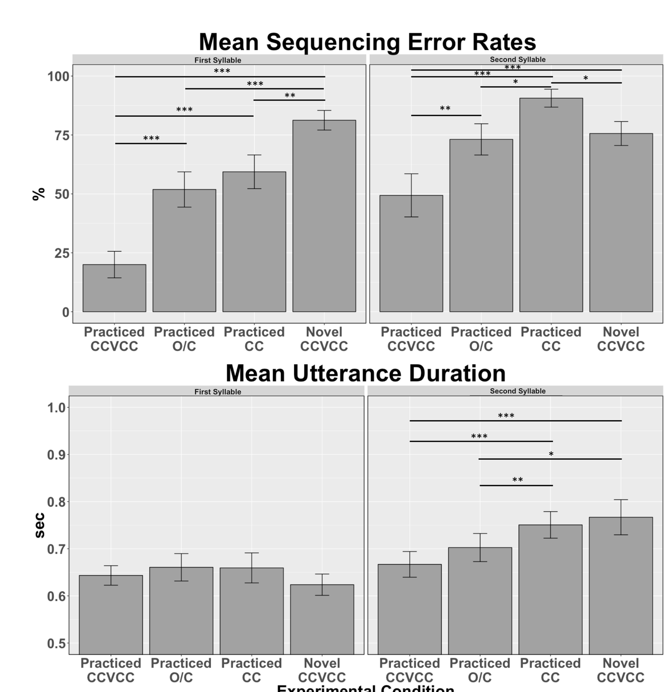
Test Phase

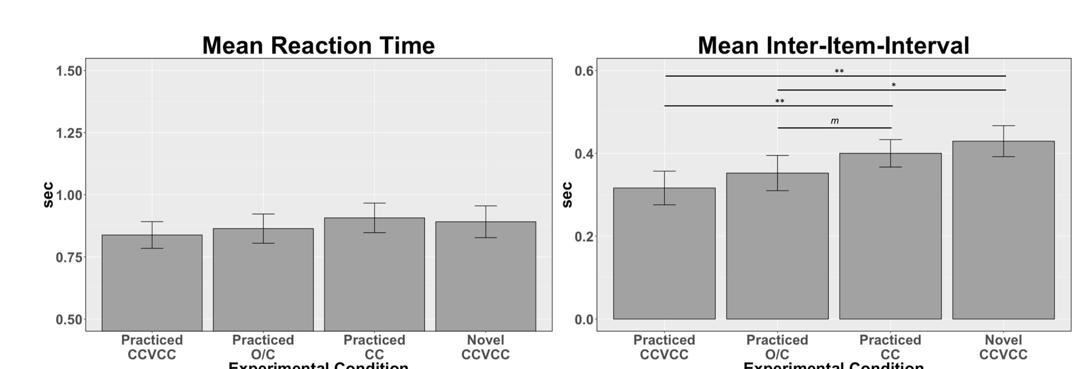
- The dependent variables were ***mean sequencing error rate, utterance duration***, ***reaction time***, and ***inter-item interval*** (i.e., time from the offset of the first syllable to the onset of the second syllable); we examined each performance measure as a function of condition and stimulus item (first vs. second syllable), since the second syllable was expected to rely more on working memory. We computed scores for each performance measure by averaging the values obtained on the first five utterances. Unlike in the training phase, even the temporal aspects of sequence generation were analyzed based on the first five utterances, regardless of whether both items were properly sequenced (i.e., it was rare for speakers to properly sequence both syllables in a given test pair).

Predicted Generalization Patterns



Actual Results

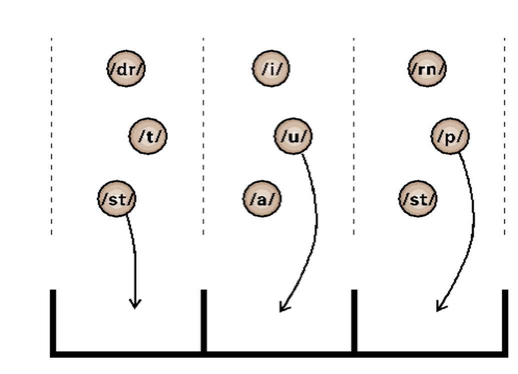




- Gains in performance speed for the second syllable only generalized to novel syllables containing previously trained clusters if the clusters were practiced in the same syllable position (onset or coda); practicing the clusters in a different position resulted in no transfer of learning.
- The accuracy results for the second syllable, however, indicated that there was no transfer of learning to novel syllables with trained clusters, and that practicing a cluster in a given context interfered with one's ability to produce it in a distinct utterance when it was embedded in a different syllable position.

CONCLUSIONS

- Collectively, the present findings suggest that, during phonological encoding, the working memory system utilizes a syllabic structural frame (onset-nucleus-coda) that codes the serial ordering of phonemes.
- On this view, adjacent tautosyllabic consonants become integrated into cohesive planning elements that are *phonologically differentiated* and "inserted" into larger syllabic frames during speech sequence assembly (cf. Shattuck-Hufnagel, 1992).



SELECTED REFERENCES

Bohland, J., 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.

Guenther, F.H. (2016). Neural Control of Speech. MIT Press, Cambridge, MA.

Segawa, J.A., Masapollo, M., Tong, M., Smith, D.J., & Guenther, F.H. (2019). Chunking of phonological units in speech sequencing. *Brain and Language*.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition*, 42, 213-259.

Buchwald, A., Gagnon, B., & Miozzo, M. (2017). Identification and remediation of phonological and motor errors in acquired sound production impairment. *Journal of Speech, Language, and Hearing Research*, 60, 1726-1738.

Funding: The research reported here was supported by the National Institute on Deafness and other Communication Disorders (R01 DC007683; F.H. Guenther, PI)..