

Temporal stability and the online assignment of hierarchical prosodic structure

Background. While hierarchical prosodic structure has long been discussed in phonological theory [e.g., 1, 2], little work has investigated whether and how listeners assign this structure to the incoming acoustic signal during spoken word recognition (SWR) [3]. This is because acoustic correlates to the prosodic structure of sublexical units (i.e., whether a target is an onset or coda) are typically multiple-use, also signaling differences in segmental identity [4]. For example, if an English plosive has a long lag voice onset time, this indicates both that the plosive is voiceless (e.g., /p, t, k/) and that it is a syllable onset [5].

However, a recent acoustic study by [6] indicates that the **temporal stability** of onset consonants could provide a reliable way to tease this information apart. Temporal stability refers to the finding that intervals between coordinated articulatory gestural landmarks are more stable than intervals between uncoordinated landmarks [7]. While most research on temporal stability measurements has been restricted to articulation, [6] found an acoustic index of temporal stability: **right-edge-to-anchor duration**, which is measured from the midpoint of the rightmost consonant (the right-edge) to the anchor (the end of the following vowel). [6] found that this interval is systematically longer in between the consonants (C) and their following vowels (V) in CVC syllable tokens than in CCVC syllable tokens in English speakers' productions.

It is not yet known whether listeners are attending to differences in right-edge-to-anchor duration to inform perception. Because right-edge-to-anchor duration is an acoustic correlate for differences in syllable structure, it is possible that listeners are using it to assign hierarchical prosodic structure to incoming lexical items during online SWR.

Methods. A lexical decision task created using PClbex [8] will be conducted with native English listeners on Prolific. The intended number of participants is 300. Following our online study, we plan to run participants in-person as a point of comparison. Target stimuli will be synthesized from natural productions of 90 CVC and 45 CCVC monosyllabic words via Praat [9]. Steady state intervals from the rightmost consonant (i.e., C_1 in C_1VC tokens and C_2 in C_1C_2VC tokens) and the vowel will be manipulated to create 3 right-edge-to-anchor duration conditions, described in (1). These manipulations, along with the 2 syllable structures (CVC vs. CCVC), result in a 2x3 factorial design. These stimuli are divided into 3 lists of target items (135 target stimuli per list) that listeners are randomly assigned to during experimentation. Along with target items, each list will include CCVC filler tokens (45 total) and nonword tokens (180 total) that have undergone the manipulations in (1). Each stimulus will be presented to listeners auditorily, and a pre-experiment test will be conducted to ensure that subjects are wearing headphones [10]. They will have 1500 ms to press a key on their keyboard to report whether what they heard was a real word or not, at which point response time (RT) measurements will be collected.

We predict that if listeners are using right-edge-to-anchor durations to assign hierarchical prosodic structure to acoustic input, then the CCVC tokens with C-like manipulations will have increased RTs. This is because the C-like duration would signal an illicit $*C_1.C_2VC$ structure (in which C_1 were extra-syllabic), which would slow processing times as listeners reanalyze the acoustic input. In contrast, if listeners are not using these durations to assign structure, then both the C-like and the CC-like stimuli will show increased RTs compared to the unaltered tokens due to both manipulations differing from what is typical for CCVC syllables. Following data collection, we will run a linear mixed model to determine the statistical significance of our results.

References. [1] Selkirk, E. (1984). On the major class features and syllable theory. In Mark Aronoff & Richard T. Oerhle (eds.), *Language Sound Structure*. MIT Press. [2] Kiparsky, P. (1979). Metrical structure assignment is cyclic. *Linguistic Inquiry*, 10(3), 421-441. [3] Frazier, L. (1987). Structure in auditory word recognition. *Cognition*, 25(1-2), 157-187. [4] Tagliapietra, L., & McQueen, J. M. (2010). What and where in speech recognition: Geminate and singletons in spoken Italian. *Journal of Memory and Language*, 63(3), 306-323. [5] Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20(3), 384-422. [6] Durvasula, K., Ruthan, M. Q., Heidenreich, S., & Lin, Y. H. (2021). Probing syllable structure through acoustic measurements: case studies on American English and Jazani Arabic. *Phonology*, 38(2), 173-202. [7] Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, 45(2-4), 140-155. [8] Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX). [9] Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5:9/10, 341-345. [10] Woods, K. J., Siegel, M. H., Traer, J., & McDermott, J. H. (2017). Headphone screening to facilitate web-based auditory experiments. *Attention, Perception, & Psychophysics*, 79, 2064-2072.

(1) Right-edge-to-anchor duration manipulations

- i) **C-like right-edge-to-anchor duration:** The right-edge-to-anchor duration will be increased to resemble that of a syllable with a simple onset (e.g., the [l] and [æ] of a “glass” token will be elongated such that its right-edge-to-anchor duration resembles an exaggerated CVC token).
- ii) **CC-like right-edge-to-anchor duration:** The right-edge-to-anchor duration will be decreased to resemble that of a syllable with a simple onset (e.g., the [l] and [æ] of a “glass” token will be shortened such that its right-edge-to-anchor duration resembles an exaggerated CCVC token).
- iii) **Unaltered right-edge-to-anchor duration:** The right-edge-to-anchor duration will remain unaltered.