

Simple acoustic measures of onset complexity

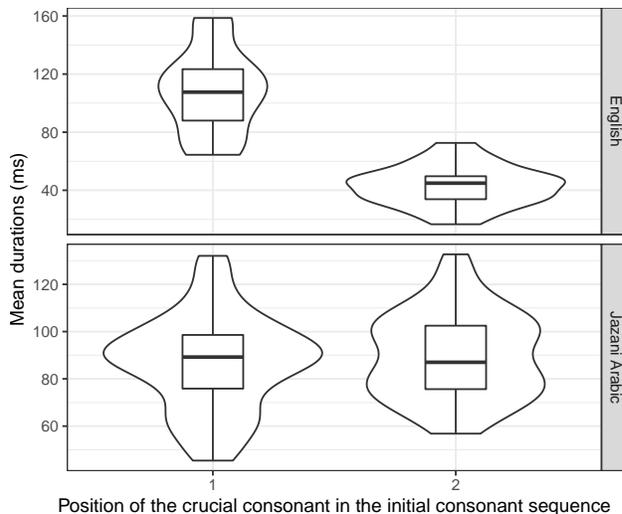
Karthik Durvasula, Michigan State University

The articulatory consequences of different onset organisations have been extensively probed since the 1980s (Browman and Goldstein 1988; Byrd 1995; Hermes et al. 2013, amongst others). In languages with complex onsets, it’s generally observed that the mean of the mid-points of the consonantal gestures of a word-initial consonant sequence (called the C-CENTRE) is temporally aligned to the end of the following vowel (called the ANCHOR), *i.e.*, the C-CENTRE is at a stable distance away from the end of the following vowel no matter how many consonants are in the sequence. In contrast, languages which allow only simplex onsets despite having word-initial consonant sequences, such as Tashlhiyt Berber (Goldstein et al. 2007; Hermes et al. 2017) and Moroccan Arabic (Shaw et al. 2009, 2011), have been observed to have a RIGHT-EDGE-TO-ANCHOR interval stability, *i.e.*, the last consonantal gesture of a word-initial consonant sequence is in a stable temporal relationship with the following vowel, and the presence of more consonants word-initially before the prevocalic consonant does not change the timing between the prevocalic consonant and the ANCHOR.

Most relevant for current purposes are two recent findings related to such temporal stability patterns: (a) Sotiropoulou et al. (2020) point out an important consequence of C-CENTRE stability in complex onset languages — a segment in the C_2 position of a $\#C_1C_2V$ sequence may be shorter than when the same segment is in the C_1 position of $\#C_1V$ sequence. A corollary they do not explore is that, in simple onset languages, a segment in the C_2 position of a $\#C_1C_2V$ sequence should be the same duration as when the same segment is in the C_1 position of $\#C_1V$ sequence. (b) Durvasula et al. (forthcoming) and Selkirk and Durvasula (2013) show that correlations between C-CENTRE/RIGHT-EDGE stability patterns and onset organisation are also observable through an analysis of *acoustic* measurements,

We combine the above two lines of research and ask if simple durational measurements observable in acoustic recordings correlate with onset complexity both in lab-based experimental data and corpus data.

Experiment 1: Experimental data We used the data published by Durvasula et al. (2021) in an Open Science Foundation repository; it includes TextGrids from production experiments on American English (a complex onset language) and Jazani Arabic (a simplex onset language). The crucial test items consisted of C_1VC and C_1C_2VC words in both languages, and the crucial word-initial consonant sequences consisted of fricatives and nasals (e.g., English: <nap, snap>; Jazani Arabic: [məf] ‘with’, [sməʕ] ‘listen’). There were 16 words for the American English experiment (8 test), and 78 words for the Jazani Arabic experiment (all test).



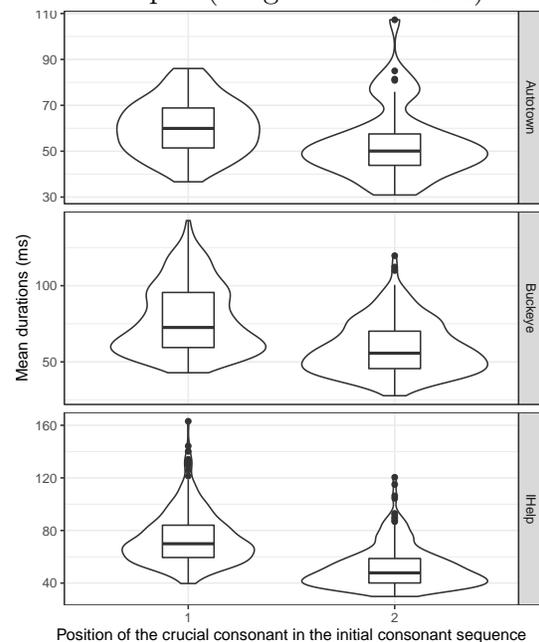
We used Praat (Boersma and Weenink 2016) scripts to extract the relevant consonant duration measurements (American English = 777 words, Jazani Arabic = 3319 words). Visual

inspection (figure above) suggests that a consonant in the second position of a word-initial consonant sequence is in fact shorter than the same consonant in a simple onset for American English but not for Jazani Arabic. We modelled the durations of the English and Jazani Arabic consonant durations using separate linear mixed effects models with a random effects structure that included by-subject, by-segment and by-corpus random intercepts and by-subject, by-segment and by-corpus random slopes for the position of the consonant. The independent variable was the position of the consonant (initial vs. peninitial). Again, we found a clear shortening of a consonant in the peninitial position of word-initial consonant sequences for English ($\hat{\beta} = -64.8$ ms, $sd = 9.73$ ms, $p(>|t|) = 0.001$), but not for Jazani Arabic ($\hat{\beta} = 3.3$ ms, $sd = 3.6$ ms, $p(>|t|) = 0.37$).

Experiment 2: Corpus data We probed if the consonant shortening was observable in naturalistic productions *without* highly controlled stimuli (which is usually the case in lab-based experiments). In short, we aimed to probe if such shortening was observable simply by comparing prevocalic consonantal durations across-the-board *without* controlling for rhymes or any other factors (word frequency, morphological complexity, ...).

For this purpose, we used high-quality recordings of conversational American English speech from three different corpora: (a) sociolinguistic interview data from the Buckeye Corpus (Pitt et al. 2005), (b) sociolinguistic interview data from the I-Help Corpus (Wagner et al. 2016), (c) oral histories documented in the Auto Town Corpus (Wagner et al. 2016).

We focussed on consonants that can appear as the second consonant in word-initial consonant sequences in American English [m, n, l, r, w, p, t, k]. We used Praat scripts to extract the relevant measurements from a total of 149,044 words from the three corpora (Buckeye = 84,046, iHelp = 56,643, Auto-town = 8,355). As with the American English results in Experiment 1, the results suggest that a consonant in the second position of a word-initial consonant sequence is in fact shorter than when it is a simplex onset (figure to the right). We modelled the durations using linear mixed effects models and found a clear shortening of about 17.5 ms of the second consonant in complex onsets ($\hat{\beta} = -17.53$ ms, $sd = 5.2$ ms, $p(>|t|) = 0.01$).



Discussion In line with our expectations, we find that there is indeed a clear effect of onset complexity on acoustic consonant durations. This result is particularly exciting because it makes it easier to collect phonetic data relevant to syllable complexity both in the lab and during fieldwork than extant methods. Furthermore, we are able to observe this effect in naturalistic corpora of acoustic measurements quite robustly even without controlling for rhyme quality and other factors — this thereby suggests, contrary to standard views, that the learner in fact has quite straightforward access the necessary phonetic information to infer syllable (onset) complexity from the acoustics.