A SIMPLE ACOUSTIC MEASURE OF ONSET COMPLEXITY

Karthik Durvasula

Michigan State University
durvasul@msu.edu

ABSTRACT

The kinematic consequences of onset complexity have been extensively probed since the 1980s — complex and simple onsets generally show different temporal stability patterns with respect to the following vowel. More recently, it has been argued that acoustic methods can also be used to establish the same dichotomy. In this paper, I show that a rather simple acoustic measure of pre-vocalic consonant duration in word-initial CV and CCV sequences can be used to tease apart complex and simple onset organisations across languages. I show this through an analysis of controlled experimental data on American English and Jazani Arabic, and through an analysis of corpus data from three different American English corpora of sociolinguistic interviews or oral histories. The results in turn suggest that a relatively coarse measure (and a relatively simple learning algorithm) may be sufficient for a learner to acquire some aspects of syllable-structure from the acoustic input.

Keywords: Onset complexity, Production experiments, Corpora, English, Jazani Arabic

1. INTRODUCTION

The kinematic consequences of different onset organisations have been extensively probed since the 1980s [1, 2, 3, amongst others]. In languages with complex onsets, it's generally observed that word-initial consonants along with the following vowel show a 'global timing stability' pattern. More specifically, 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 (see Figure 1, left). This pattern of c-centre-to-anchor interval stability has since been replicated for American English [4], and has also been observed in a variety of languages with complex onsets: Georgian [5], Italian [2], Polish [6], Romanian [7], and Spanish [8]. In contrast, word-initial consonant sequences in languages which allow at most a single consonant (simple onsets languages) have been observed to show a local timing stability' pattern. More specifically, word-initial consonant sequences have a right-edge-to-anchor interval stability, wherein 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 (see Figure 1, right). This pattern of right-edge-to-anchor interval stability has been observed in Tashlhiyt Berber [5, 6], Moroccan Arabic [9, 10] and Jazani Arabic [11],

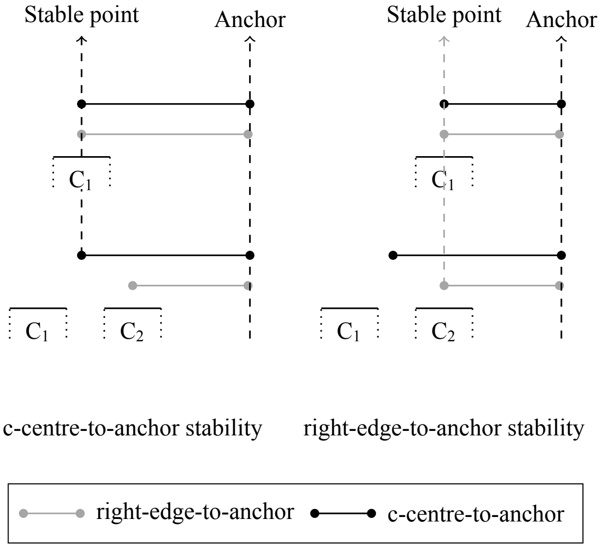

Figure 1: Schematic representations of c-centre-to-anchor stability patterns (left) and right-edge-to-anchor stability patterns (right). The x-axis in the figure represents time. The anchor marks the end of the following vowel, and C_1-C_2 represent word-initial consonants.

There are however a few results that contradict the above linking hypothesis between temporal stability patterns and onset complexity in a language. There are arguments in the phonological literature that suggest that word-initial consonant sequences in Hebrew [12], French [13] and German [14] form complex onsets. However, the three languages have been observed to show a right-edge alignment, at least for some consonant sequences [15, 16, 17]. Relatedly, the c-centre-to-anchor stability pattern was observed for only some of the consonant clusters studied in Romanian [18] and Polish [6], though all the clusters studied were previously argued to be complex onsets.

While the observations might be at first blush seem problematic for the linking hypothesis discussed above, Mücke, Hermes, and Tilsen [19] recently argued that the patterns observed in the languages are consistent with a complex onset organisation, and previous research likely misinterpreted the relevant articulatory data. More specifically, they suggest that explicitly modelling the speaker-specific coupling strength between gestures and speaker-specific biomechanical interactions between articulators allows us to still understand the patterns in such languages as stemming from a c-centre organisation. They show this is possible at least for German. A second possibility to account for the discordant results is discussed by Durvasula, Ruthan, Heidenreich, and Lin [11], who suggest based on their results that it is possible that the relevant articulatory data is in fact indirect evidence of the the stability patterns present in the acoustics, and that there might be more stability for the c-centre-to-anchor than for the right-edge-to-anchor for the three languages, when the intervals are extracted from acoustic measurements. Both of the above suggestions raise the possibility that the articulatory stability patterns observed in the three languages may not be counterexamples to the linking hypothesis after all. A third possibility raised by Sotiropoulou, Gibson, and Gafos [8] is that the temporal stability pattern is only a part of the gamut of correlates that distinguish global timing stability in complex onset languages from local timing stability in simple onset languages, and really the distinction between the two is simultaneously expressed over a set of different phonetic parameters rather than just through a single measure such as c-centre or right-edge stability. In this paper, I take inspiration from the last two possibilities.

Most relevant for current purposes are two recent findings: (a) Sotiropoulou, Gibson, and Gafos [8] point out an important correlate of global timing 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, but is necessary to complete their argument, is the following: 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, Ruthan, Heidenreich, and Lin [11] and Durvasula and Selkirk [20] show that correlations between c-centre/right-edge stability patterns and onset organisation are also observable through an analysis of *acoustic* measurements. I combine the above two lines of research and ask if the simple durational measurement of pre-vocalic consonant duration observable in acoustic recordings correlates with onset complexity both in a lab-based production experiment and in three different corpora of American English.

2. EXPERIMENT 1: PRODUCTION EXPERIMENT

I used the data published by [21] in an Open Science Foundation repository; it includes TextGrids from production experiments on 10 American English speakers (a complex onset language) and 7 Jazani Arabic speakers (a simple 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: [məʕ] ‘with’, [sməʕ] ‘listen’). There were 16 words for the American English experiment (8 test), and 78 words for the Jazani Arabic experiment (all test). The actual counts of the consonantal positions in each experiment are shown in Table 1.

Corpus	C_1	C_2
English	389	388
Jazani Arabic	1850	1469

Table 1: Counts of the test consonants in relevant positions (underlined) in $\#\underline{C_1}V\ldots$ and $\#C_2\underline{C_2}V\ldots$ words in each of the three corpora.

Praat [**?**] scripts were used to extract the relevant consonant duration measurements. Visual inspection suggested 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 (see Figure 2).

The English and Jazani Arabic consonant durations were modelled using the programming language R [22]. Separate linear mixed effects models were fitted using the R packaes `lme4` and `lmerTest` [23, 24] for each language with a random effects structure that included by-subject, and by-segment random intercepts and by-subject, by-segment and by-corpus random slopes for the position of the consonant. The independent variable

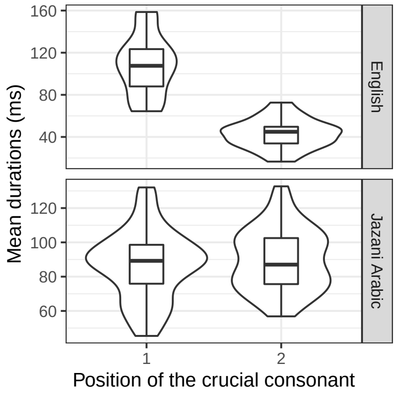

Figure 2: Durations of pre-vocalic consonants in initial and second positions of word-initial consonant sequences in the experimental data; English (top) vs. Jazani Arabic (bottom)

was Consonant Position (initial vs. second). As with the initial visual inspection, there was a statistically clear shortening of a consonant in the second position of word-initial consonant sequences compared to the same consonants in initial position for English ($\hat{\beta}$ = -64.8 ms; 95% CIs = -83.7 – -45.5 ms; see Table 2), but not for Jazani Arabic ($\hat{\beta}$ = 3.3 ms; 95% CIs = -4.0 – 10.6 ms; see Table 3). [Note: throughout, the phrase "statistically clear", and variants, are used instead of "statistically significant" on the recommendation of [25].]

	Est.	df	t	Pr(>|t|)
Interval	107.86	9.20	11.19	< 0.0001
Position: 2	-64.78	4.65	-6.64	0.002

Table 2: Linear mixed-effects model for the English consonant duration (random effects structure: by-subject, by-segment random intercepts and random slopes of Consonant Position)

	Est.	df	t	Pr(>|t|)
Interval	86.66	9.63	12.22	< 0.0001
Position: 2	3.32	12.60	0.92	0.37

Table 3: Linear mixed-effects model for the Jazani Arabic consonant duration (random effects structure: by-subject, by-segment random intercepts and random slopes of Consonant Position)

3. EXPERIMENT 2: CORPUS DATA

Subsequent to the analysis presented above, I probed if the consonant shortening was observable in naturalistic productions *without* highly controlled stimuli that are usually the case in laboratory experiments. In short, I 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, I used high-quality recordings of conversational American English speech from three different corpora: (a) **The Buckeye corpus**: conversational speech from 40 speakers [26]. The speakers were all natives of Central Ohio, and were recorded in Columbus, Ohio before Spring 2000. The sample design was stratified for age (under thirty and over forty) and sex. The recordings were hand-transcribed, and both the recordings and the transcriptions are available at https://buckeyecorpus.osu.edu. (b) **The Influence of Higher Education on Local Phonology (I-Help) corpus**: 60 sociolinguistic interviews conducted in 2014 [27]. The speakers were all natives of the Greater Lansing area of Michigan. In this article, I report the the results of an analysis of 44 speakers (29 self-reported female, and 15 self-reported male) that were annotated and I was given access to by the authors. The interviews were transcribed and time-aligned by the original authors in ELAN [28]. (c) **The Auto Town Corpus**: 21 oral histories collected in the mid-1990s and 2000s [27]. The speakers (9 self-reported women, 12 self-reported male) who varied in ages (date of birth range: 1907-1971) were all former auto plant workers of the Local 602, Fisher Body and Diamond REO assembly plants in Lansing, Michigan. The interviews were again transcribed and time-aligned by the original authors in ELAN.

The focus was 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]. All the words where the relevant consonants were the first consonant in #CV... sequences and the second consonant in #CCV... were identified. No other restrictions were imposed to narrow down the word-lists. Praat scripts were used to extract the relevant measurements from a total of 149,044 words from the three corpora (Buckeye = 84,046, I-Help = 56,643, Autotown = 8,355). The actual counts of the relevant consonantal positions in each corpus are shown in Table 4.

Corpus	C_1	C_2
Autotown	7111	1244
Buckeye	72488	11558
I-Help	48165	8478

Table 4: Counts of the test consonants in relevant positions (underlined) in $\#\underline{C_1}V...$ and $\#C_2\underline{C_2}V...$ words in each of the three corpora.

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 simple onset (see Figure 3). The durations were modelled using a linear mixed effects model 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. In line with the visual inspection, a statistically clear shortening ($\hat{\beta}$ = -17.5 ms; 95% CIs = -28.2 – -6.5 ms) was observed for the second consonant in complex onsets compared to the same consonants in initial position (see Table 5).

	Est.	df	t	Pr(>|t|)
Interval	71.89	8.07	10.57	< 0.0001
Position: 2	-17.53	7.99	-3.35	0.01

Table 5: Linear mixed-effects model for the consonant duration (random effects structure: by-subject, by-segment and by-corpus random intercepts and random slopes of Consonant Position)

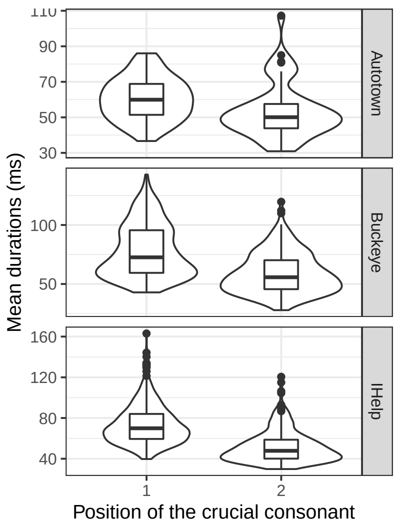

Figure 3: Durations of the consonants in relevant positions (underlined) in #$\underline{C_1}$V... and #$C_2\underline{C_2}$V... words in the three corpora.

4. DISCUSSION

A clear effect of onset complexity on acoustic consonant durations was found. There was no statistically clear difference in the durations of pre-vocalic consonants in word-initial CV and CCV words in Jazani Arabic, but there was a statistically clear difference in the durations of similar words for American English, both in controlled production data collected in a laboratory and in much more naturalistic sociolinguistic interviews or oral histories. Furthermore, the 95% confidence intervals for the controlled production experiment were quite far away from zero (-83.7 – -45.5 ms), while the same for the controlled Jazani Arabic production experiment included a very small range close to zero (-4.0 – 10.6 ms). This suggests that the differences observed in the American English cases are quite meaningful, and substantially different in character from Jazani Arabic.

The results therefore suggest that the acoustic shortening of the pre-vocalic consonant duration could indeed be a reliable cue to differentiate consonant sequences that form complex onsets from those that don't. We echo the important point made in Durvasula, Ruthan, Heidenreich, and Lin [11] that the presence of such syllable-structure signatures in the acoustic recordings makes it easier to collect phonetic data relevant to syllable complexity both in the lab and during fieldwork than extant methods.

The results also provide further support for claims in the phonological literature that word-edge (consonant) sequences do not constitute direct evidence of syllable structure in a language [29], and therefore highlight the need for clear linking hypotheses connecting abstract structure to phonological patterns or phonetic data (such as the one explored in the current paper) in order to identify the abstract structure.

Finally, this effect was observed in naturalistic corpora of acoustic measurements quite robustly. Given the uncontrolled nature of the stimuli, it is of course to be expected that the unstandardised effect size in the corpus analysis was much smaller than the same in the controlled production experiment for American English. Furthermore, the corpus results are particularly exciting because they suggest, contrary to standard views, that the acoustic input not only has information about some aspects of syllable-structure, but that relatively coarse data is sufficient to observe this pattern, which thereby suggests that an extremely simple learning algorithm that keeps track of pre-vocalic consonant durations is potentially sufficient for a learner to acquire some aspects of syllable-structure directly from the acoustic input they receive.

5. REFERENCES

[1] C. P. Browman and L. H. Goldstein, "Some notes on syllable structure in articulatory phonology," *Phonetica*, vol. 45, pp. 140–155, 1988.

[2] A. Hermes, D. Mücke, and M. Grice, "Gestural coordination of Italian word-initial clusters: The case of 'impure s'," *Phonology*, vol. 30, no. 1, pp. 1–25, 2013.

[3] D. Byrd, “C-centers revisited,” *Phonetica*, vol. 52, pp. 263–282, 1995.

[4] S. Marin and M. Pouplier, “Temporal organization of complex onsets and codas in american english: Testing the predictions of a gestural coupling model,” *Motor Control*, vol. 14, no. 3, pp. 380–407, 2010.

[5] L. H. Goldstein, I. Chitoran, and E. Selkirk, “Syllable structure as coupled oscilator models: Evidence from georgian vs. tashlhiyt berber,” *Proceedings of the 16th International Congress of Phonetic Sciences*, pp. 241–244, 2007.

[6] A. Hermes, D. Mücke, and B. Auris, “The variability of syllable patterns in tashlhiyt berber and polish,” *Journal of Phonetics*, vol. 64, pp. 127–144, 2017.

[7] S. Marin and M. Pouplier, “Articulatory synergies in the temporal organization of liquid clusters in romanian,” *Journal of Phonetics*, vol. 42, pp. 24–36, 2014.

[8] S. Sotiropoulou, M. Gibson, and A. Gafos, “Global organization in spanish onsets,” *Journal of Phonetics*, vol. 82, p. 100995, 2020. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0095447020300863

[9] J. A. Shaw, A. I. Gafos, P. Hoole, and C. Zeroual, “Syllabification in moroccan arabic: evidence from patterns of temporal stability in articulation,” *Phonology*, vol. 26, no. 1, pp. 187–215, 2009.

[10] ——, “Dynamic invariance in the phonetic expression of syllable structure: a case study of moroccan arabic consonant clusters,” *Phonology*, vol. 28, no. 3, pp. 455–490, 2011.

[11] K. Durvasula, M. Q. Ruthan, S. Heidenreich, and Y.-H. Lin, “Probing syllable structure through acoustic measurements: case studies on american english and jazani arabic,” *Phonology*, vol. 38, no. 2, p. 173–202, 2021.

[12] S. Bolozky, “Israeli hebrew phonology,” in *Phonologies of Asia and Africa (including the Caucasus). Vol. 1*, A. S. Kaye, Ed. Eisenbrauns, 1997, pp. 287–311.

[13] F. Dell, “Consonant clusters and phonological syllables in french,” *Lingua*, vol. 95, no. 1, pp. 5 – 26, 1995, french Phonology. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0024384195900993

[14] R. Wiese, *The Phonology of German*. Clarendon, 1996.

[15] M. Pouplier, “The gestural approach to syllable structure: universal, language- and cluster-specific aspects,” pp. 63–96, 2012.

[16] J. Brunner, C. Geng, S. Sotiropoulou, and A. Gafos, “Timing of german onset and word boundary clusters,” *Laboratory Phonology*, vol. 5, pp. 403–454, 2014.

[17] S. Tilsen, D. Zec, C. Bjorndahl, B. Butler, M.-J. L’Esperance, A. Fisher, L. Heimisdottir, M. Renwick, and C. Sanker, “A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters,” *Cornell Working Papers in Phonetics and Phonology 2012*, pp. 51–81, 2012.

[18] S. Marin, “The temporal organization of complex onsets and codas in romanian: A gestural approach,” *Journal of Phonetics*, vol. 41, no. 3, pp. 211–227, 2013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0095447013000168

[19] D. Mücke, A. Hermes, and S. Tilsen, “Incongruencies between phonological theory and phonetic measurement,” *Phonology*, vol. 37, no. 1, pp. 133–170, 2020.

[20] E. Selkirk and K. Durvasula, “Acoustic correlates of consonant gesture timing in english,” 2013, paper presented at the 166th Meeting of the Acoustical Society of America, San Francisco, USA.

[21] K. Durvasula, M. Ruthan, S. Heidenreich, and Y.-H. Lin, “Osf repository for *Probing Syllable Structure Through Acoustic Measurements: Case-studies on American English and Jazani Arabic*,” 2021.

[22] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: http://www.R-project.org

[23] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting linear mixed-effects models using lme4,” *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[24] A. Kuznetsova, P. B. Brockhoff, and R. H. B. Christensen, “lmerTest package: Tests in linear mixed effects models,” *Journal of Statistical Software*, vol. 82, no. 13, pp. 1–26, 2017.

[25] J. Dushoff, M. P. Kain, and B. M. Bolker, “I can see clearly now: Reinterpreting statistical significance,” *Methods in Ecology and Evolution*, vol. 10, no. 6, pp. 756–759, 2019. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13159

[26] M. A. Pitt, K. Johnson, E. Hume, S. Kiesling, and W. Raymond, “The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability,” *Speech Communication*, vol. 45, no. 1, pp. 89–95, 2005.

[27] S. E. Wagner, A. Mason, M. Nesbitt, E. Pevan, and M. Savage, “Reversal and re-organization of the northern cities shift in michigan,” *University of Pennsylvania Working Papers in Linguistics*, vol. 22, no. 2, pp. 171–179, 2016. [Online]. Available: https://repository.upenn.edu/pwpl/vol22/iss2/19/

[28] P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes, “Elan: A professional framework for multimodality research,” in *5th international conference on language resources and evaluation (LREC 2006)*, 2006, pp. 1556–1559.

[29] T. Borowsky, “Topics in the lexical phonology of English,” Ph.D. Dissertation, University of Massachusetts, Amherst, MA, USA, 1986. [Online]. Available: https://scholarworks.umass.edu/dissertations/AAI8701140/