DISSECTING THE CONSONANT DURATION RATIO

Stavroula Sotiropoulou[1], Erika Pillmeier[1], Argyro Katsika[2] and Adamantios Gafos[1, 2]

[1]University of Potsdam, Potsdam, Germany, and [2]Haskins Laboratories, New Haven, CT, USA
stsotiro@uni-potsdam.de; erika.pillmeier@uni-potsdam.de; argyro.katsika@yale.edu; gafos@uni-potsdam.de

ABSTRACT

It has been argued that the duration ratio (DR) of two consonants in C1C2 serves as a diagnostic of syllabification: if greater than 1, then C1C2 is a syllable onset; if (approximately) 1, C1C2 is a coda–onset sequence. If valid, this diagnostic would provide a straightforward way of assessing syllabic organization. We examined the validity of the DR using the Wisconsin X-Ray Microbeam corpus, which due to its size permits a robust test. We also informed our evaluation by another archive of kinematics and acoustics, with finer comparisons. These corpora extend an evaluation of the DR to word-medial clusters, in addition to initial and cross-word contexts wherein the DR has been studied before. Our results indicate that duration ratio patterns previously thought to be related to syllabification derive from levels of representation or prosody which are not specifically syllabic. These patterns are conditioned by prosodic effects and manner contrasts.

Keywords: duration ratio hypothesis; syllabification indexes; prosodic structure; π-gesture.

1. INTRODUCTION

It has been argued that the duration ratio (DR) of the two consonants in a C1C2 cluster serves as a diagnostic of syllabification: if greater than 1, then the cluster C1C2 is parsed as a (complex) syllable onset; if approximately equal to 1, the cluster is a sequence of a coda consonant (C1) followed by an onset consonant (C2), split across two syllables [5] [9] [10]. We refer to this diagnostic as the Duration Ratio Hypothesis (DRH). If valid, this diagnostic would provide a straightforward way of assessing syllabic organization in clusters, obviating more involved approaches to the problem (e.g. see [11]).

We evaluated the validity of the DR using Wisconsin's X-Ray Microbeam corpus [12], which due to its size (57 speakers, Am. English, kinematics and acoustics) and variety of data (isolated words, phrases, but also whole paragraphs) permitted a robust test. We also informed our assessment by the Potsdam archive of Electromagnetic Articulometry (EMA), which provided finer control on comparisons but more limited data. Similarly to the Wisconsin X-Ray Microbeam corpus, the Potsdam archive was analyzed both in terms of acoustics and kinematics. It also allowed us to extend for the first time an evaluation of the DR to word-medial clusters, in addition to word-initial and cross-word contexts wherein the DR has been studied before. According to the DRH, word-medial legal onset clusters should behave like word-initial clusters because both are syllable onsets (regardless of word position).

2. ARTICULATORY DATA AND RESULTS

2.1. Wisconsin X-Ray Microbeam archive

We used the Wisconsin X-Ray Microbeam (XRMB hereafter) corpus to quantify duration ratios in the following clusters: /sn, sm, fl, gl, tw/ in syllable-initial and cross-word position (e.g., *smooth* vs. *this march*). There were 24 to 35 speakers per cluster analyzed. This was the maximal number of clusters (and speakers per cluster) that could be obtained, while in parallel allowing us to systematically investigate the effect of prosodic position (utterance-initial, utterance-medial). Specifically, only clusters that appeared both in syllable-initial position and in cross-word position were used (see Table 1). All syllable-initial clusters were word-initial as well, except /fl/ and /tw/ that also had word-medial counterparts (marked with an asterisk in Table 1) that were included in one set of analyses (further described below). The clusters were further divided with respect to their position in the utterance into utterance-initial and utterance-medial. Due to the nature of the archive, the design of the stimuli was not balanced. Only the clusters /fl/ and /tw/ appeared in both utterance positions (initial, medial). An additional requirement was that the clusters were composed of heterorganic segments. This was to aid in measurements, as homorganic segments cannot be accurately measured in kinematics since they use the same articulator. The only exception to this requirement was /sn/, because frication in the acoustics tends to correspond well to the gestural plateau of the fricative (a detailed description of /sn/ delimitation follows). The heterorganic /sw/ cluster was in principle also available, but it was excluded from the analysis, because its articulatory components could not be reliably identified in the given context.

The clusters included in the analysis are summarized in Table 1, along with a list of their properties, i.e., location of the coil (LL stands for lower lip, TT for tongue tip, and TB for tongue body), the items used for their elicitation per word position (syllable-initial and cross-word), their position in the utterance (initial or medial), the number of the correspondent task as originally defined in the corpus' description, and the number of tokens used.

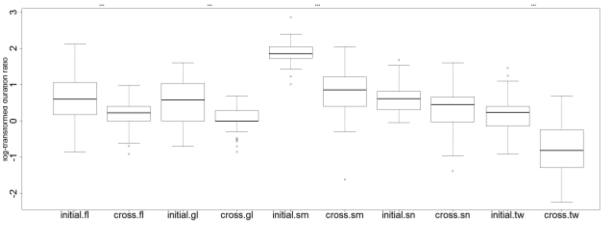
Table 1: XRMB items and respective coils measured along with archive Task number and number of repetitions.

Clusters	Coils	Syllable-initial	Cross-word	Utterance position	Task #	N
/sn/	/s/: TT	snow		medial	12	43
	/n/: TT		this nothing	initial	76	27
/sm/	/s/: TT	smooth		initial	35	25
	/m/: LL		this march	medial	78, 79	39
/fl/	/f/: LL	flip,		initial	65	33
	/l/: TT	butterfly*		medial	43	26
			off like,	medial	36	32
			lifelong	medial	81	32
/gl/	/g/: TB	glowing		initial	4	32
	/l/: TT		rag like	medial	40, 68	69
/tw/	/t/: TT	twice,		initial	11, 12	35
	/w/: LL	between*		initial, medial	19, 28, 40, 74, 86, 101	150
			light wax,	initial	91	23
			coat wax	initial	95	25

Consonant plateaus were measured via automatic labelling using Mview in MATLAB (designed by Mark Tiede of Haskins Laboratories). Plateau duration was defined as the interval from plateau onset (achievement of target for constriction) to plateau offset (release of constriction) (see Figure 3, which illustrates the plateaus of two consecutive consonants from the Potsdam archive). In the case of /sn/, a single TT plateau corresponded to both /s/ and /n/. The onset of the plateau corresponded to the constriction achievement for /s/ and coincided with the onset of high frequency noise for the fricative. The offset of the plateau defined the release of the /n/ constriction, and co-occurred with the offset of the faint formant structure for the nasal (and consequently with the onset of a clear formant structure for the subsequent vowel as well). The articulatory boundary between /s/ and /n/ was set to the offset of the high frequency noise for /s/ (which also noted the onset of the formant transition to /n/).

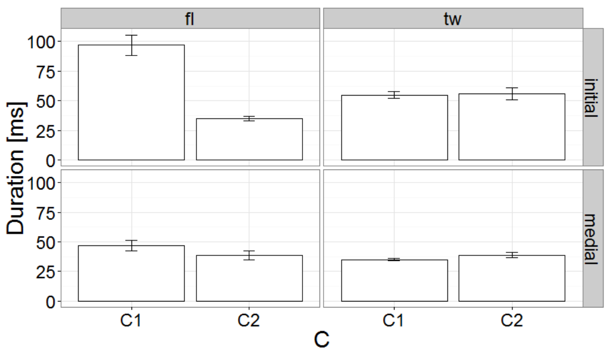
The statistical analysis was performed with R version 2.14.1 [8] using ANOVA along with t-tests with a 0.05 significance level. When comparing the different conditions (initial, cross) within cluster, t-tests reveal significant results for all clusters ($p < .001$) apart from /sn/ cluster. The results show that when the prosody is not controlled, lumping utterance-initial and medial position together, the DR behaves as expected across all conditions, except for cluster /sn/, where the duration ratio between /s/ and /n/ remains the same regardless of word position (Figure 1).

Figure 1: Log-transformed duration ratios per cluster for initial and cross-word clusters. "***" represents a significant difference between indicated conditions at $p < .001$.



To control prosody, we examined /fl/ and /tw/, the only clusters available in both initial and medial utterance positions in the XRMB archive. In this analysis, the word-medial /fl/ and /tw/ were used, since these were the only utterance-medial instances of /fl/ and /tw/. The results showed that only C1 in the fricative-liquid (i.e., /fl/) clusters seems to lengthen enough to exhibit the DRH predicted pattern (Figure 2, left). A repeated measures ANOVA with C1 plateau duration as the dependent variable and cluster and utterance position as factors yielded a significant main effect of utterance position when the data consisted of only /fl/ and /tw/ clusters ($F(1, 205= 15.54$, $p= .0001$)

Figure 2: Plateau duration (in ms) for /fl/ on the left and /tw/ on the right; comparing utterance-initial vs. utterance-medial positions.



2.2. Potsdam EMA archive

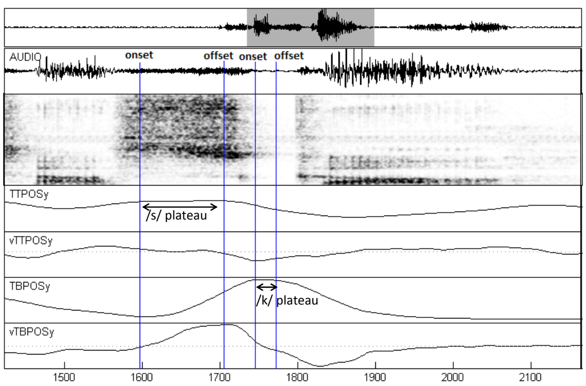
Kinematic data were collected using a Wave NDI electromagnetic articulometer at a sample rate of 400 Hz. Concurrent audio data were captured at a sample rate of 25.6 kHz using a National Instruments Modular data acquisition system (DAQ). The subset of the Potsdam archive corpus analyzed here consists of the following clusters: /sp, sk, br, kl/. The target sequence was always in accented position in the carrier sentence "I saw '(target sequence)' on the sign". Eight repetitions of each test sentence were acquired. The data of two American English speakers (male) were analyzed. Table 2 organizes the clusters used in this analysis per coil (lower lip (LL), tongue tip (TT) and tongue body (TB)), position in the word (word-initial, word-medial and cross-word) and number of tokens analyzed.

Table 2: Potsdam EMA Archive items and coils.

Cluster	Coils	Word-initial	Word-medial	Cross-word	N
/sk/	/s/: TT /k/: TB	skate, scheme, scull	gaskeep, discover	less kate, less keep, less cull	64
/sp/	/s/: TT /p/: LL	spend, speech, spike	suspend, aspeed, despite	less pend, less peach, less pike	72
/kl/	/k/: TB /l/: TT	climb, clip, clean	decline, eclipse	rack lime, rack lip, rack lean	64
/br/	/b/: LL /r/: TT	brain, breed, brush	abrade, gabreed, abrupt	gab rain, gab read, gab rush	72

An example of a measurement of /sk/ in "scull" is illustrated in Figure 3. The /s/ plateau is delimited by the onset and offset of the vertical displacement of the tongue tip (TTPOSy). The /k/ plateau is delimited by the onset and offset of the vertical displacement of the tongue body (TBPOSy).

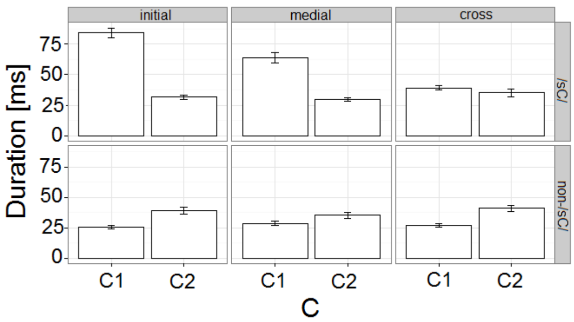
Figure 3: An instance of /sk/ in "scull" in mtnew. Panels from top to bottom: audio signal for the entire task, audio signal for a zoomed in area followed by its spectrogram, vertical displacement of the tongue tip (TTPOSy) followed by its velocity signal (vTTPOSy), vertical displacement of the tongue body (TBPOSy) followed by its velocity signal (vTBPOSy).



Repeated measures ANOVA was conducted for sC and non-sC clusters separately. Additional factors were cluster and condition (word position) and their interaction. Item was included as error term. There were no significant results for the non-sC clusters. For the sC clusters, there was a significant main effect of condition ($F(2, 11)= 8.639$, $p< .01$). The t-tests on condition showed no significant difference between word-initial and word-medial position. However, the cross-word position is significantly different from the other two ($p< .001$). sC clusters show the expected pattern in terms of DRH. Specifically, as Figure 4 (top row) shows, C1 (/s/) is longer in word-initial and word-medial positions (i.e., as syllable onset) than in cross-word ones (i.e., as syllable coda) ($p< .001$ and $p< .03$, respectively). C2 remains stable in its duration. Therefore, any change in DR is the result of C1's length changes.

For the non-sC clusters, the DR does not show the expected pattern. On the contrary, C1 and C2 duration remain stable across the three positions (Figure 4, bottom row).

Figure 4: sC clusters behave differently from non-sC clusters. Plateau durations per cluster class and position in word (Top row: sC; Bottom row: non-sC).

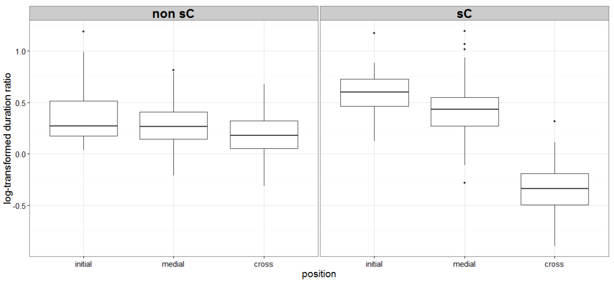


3. ACOUSTIC DATA AND RESULTS

The acoustics were extracted from the Potsdam EMA archive. The subjects and clusters were the same as in the EMA data. A linear mixed effects model was fitted for each group of clusters (sC, non-sC) with duration ratio as the dependent variable and "word position" as fixed effect. Due to limited subjects (two), the model could not estimate the random effects variance and thus "subject" was also treated as a fixed effect to account for data variance. "Item" was treated as a random factor.

The results converge with the articulatory data. For the non-sC clusters, there were no significant differences across positions. For the sC clusters, a significant difference is found between the word-initial and cross-word position ($p< .001$) and between the word-medial and cross-word position (p< .001). As can be seen in Figure 5, the DRH seems to hold true for the sC clusters (/sp, sk/), but for the non-sC clusters (/br, kl/) there are no reliable DR differences.

Figure 5: Duration ratio across positions for non-sC clusters (left) and sC clusters (right).



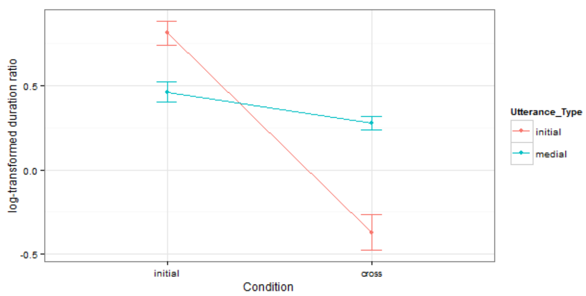
4. WHENCE THESE PATTERNS?

Our results indicate that the DR is not an index to syllabification, but rather a by-product of prosodic modulations. As illustrated in Figures 2 and 4, the first consonant in a fricative (/s/ or /f/)-initial cluster changes more across word positions than a non-fricative first consonant, and this property is what gives rise to the observed patterning of the DR. There is substantial converging evidence for such an asymmetry in independent work. Fricatives undergo position- and stress-related lengthening [3]. In contrast, stops in strong prosodic positions (initial, final or prominent) are not subject to prosody-driven effects to the same degree as their fricative counterparts. For instance, in her examination of phrase-final stops [1] and fricatives [2], Berkovits found greater boundary lengthening on the fricatives than on the stop closures [cf. 6, 7]. Considering the π-gesture model [4], the segmental composition of clusters stretches to match the activation level of the suprasyllabic π-gesture. While all consonants are equally acted upon by the π-gesture, only consonant manners that have the ability to stretch more are likely to display the ratio.

Furthermore, when the convergence of prosodic conditions leading to the expected patterning of the DR is not met, even fricative-stop clusters fail to show the pattern. For example, according to the expected pattern, DRs between word-initial and cross-word sC should be different with the former's DR greater than that of the second. However, a t-test comparison between word-initial /sn/ taken from an utterance-medial position and cross-word /sn/ shows no significant difference in DR (p= 0.13). The lengthening of /s/ in that word-initial position is not as strong as would be required to raise the DR to a value different from the cross-word condition.

In order to examine whether the effect of syllabification on DR depends on utterance position, a repeated measures ANOVA was conducted. DR was the dependent variable and utterance position, C1 manner, cluster class, condition (word position) and an interaction between utterance position and condition were included as factors. Indeed, across our corpus in addition to main effects of C1 manner ($F(1, 3)= 30.2$, $p= .0119$) and word position ($F(1, 3)= 12.337$, $p= .0391$), we also find an interaction between utterance position and word position ($F(1, 3)= 18.963$, $p= .0224$). As shown in Figure 6, for utterance-initial position the DR changes depending on the cluster's syllabification, but for utterance-medial position the DR remains relatively stable regardless of syllabification.

Figure 6: Interaction between utterance position (initial, medial) and condition (word-initial, cross-word) over all clusters.



In sum, our results indicate that the duration ratio in clusters is modulated by asymmetries in prosodic stretching between stops and fricatives and that the factors driving these modulations do not make reference to syllables.

5. CONCLUSION

We assessed the validity of the duration ratio as an index of syllabic organization in both acoustic (as done before) and articulatory data. We also extended the evaluation of the duration ratio to word-medial clusters in addition to word-initial and cross-word clusters studied before. Our results indicate that duration ratio patterns are not related to syllabification, as previously thought. Instead, they stem from the interaction of prosodic structure with segmental dimensions that determine the expandability of the segment, such as manner of articulation.

6. ACKNOWLEDGEMENTS

Stavroula Sotiropoulou and Adamantios I. Gafos gratefully acknowledge support by ERC AdG 2494406.

7. REFERENCES

[1] Berkovits, R. 1993a. Utterance-final lengthening and the duration of final-stop closures. *Journal of Phonetics* 21, 476-489.

[2] Berkovits, R. 1993b. Progressive utterance-final lengthening in syllables with final fricatives. *Language and Speech* 36, 89-98.

[3] Byrd, D., Choi, S. 2010. At the juncture of prosody, phonology, and phonetics: the interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Laboratory Phonology* 10. Mouton de Gruyter.

[4] Byrd, D. & Saltzman, E. 2003. The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics* 31, 149-180.

[5] Christie, W.M. 1977. Some multiple cues for juncture in English. *General Linguistics* 17, 212-222.

[6] Klatt, D.H. 1975. Vowel lengthening is systematically determined in a connected discourse. *Journal of Phonetics* 3, 129-140.

[7] Oller, D.K. 1973. The effect of position in utterance on speech segment duration in English. *J. Acoust. Soc. Am*. 54, 1235-47.

[8] R Development Core Team. 2011. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org/

[9] Redford, M. A. 2003. Cognitive template for a phonetic correlate of syllable structure. In: Solé, M. J., Recasens, D. & Romero J. (eds.), *ICPhS 15*. Casual Productions, 2261-2264.

[10] Redford, M.A., Randall, P. 2005. The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics* 33, 27-46.

[11] Shaw, J. & Gafos, A. 2010. Quantitative evaluation of competing syllable parses. In: Heinz, J., Cahill, L. and Wicentowski, R. (eds), *ACL SIGMORPHON-11*, 54–62.

[12] Westbury, J.R. 1994. *X-ray microbeam speech production database user's handbook*. University of Wisconsin.